# INFORSID 2024

42e édition

iNforsiD

## 28-31 mai Loria - Nancy

INFormatique des ORganisations
et Systèmes d'Information et de Décision

Président du Comité de Programme
**Max Chevalier**
IRIT, Université Paul Sabatier

Président du Comité d'Organisation
**Khalid Benali**
Loria, Université de Lorraine

inforsid2024.sciencesconf.org

Loria cnrs UNIVERSITÉ DE LORRAINE Inria métropole GrandNancy PÔLE AM2I

# Préface

Je suis très heureux de vous présenter les actes du 42ème congrès INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision) qui est organisé, cette année, dans la belle ville de Nancy.

Le congrès INFORSID constitue un lieu d'échange privilégié entre chercheurs et praticiens pour identifier et explorer les problématiques, les opportunités et les solutions que les Systèmes d'Information (SI) apportent ou absorbent.

C'est aussi l'occasion de partager et de diffuser les expériences de mise en œuvre des méthodes, modèles, outils et solutions liés aux nouvelles technologies. Les SI sont omniprésents au sens où aucun domaine d'activité n'y échappe. Aussi, dans certains secteurs (santé, finance, agriculture, environnement, culture, logistique, etc.), les SI posent des défis qui sont propres à leurs secteurs, ouvrant ainsi des champs d'étude spécifiques.

Le congrès INFORSID est également un formidable vecteur d'intégration des collègues, des professionnels mais également des doctorants travaillant dans des thématiques liées aux SI. Je me souviens qu'une des premières conférences où j'ai soumis et présenté un article était INFORSID (en 2001 😊).

C'est notamment pour cela que je suis très heureux que, cette année encore, le **forum Jeunes Chercheuses Jeunes Chercheurs JCJC** soit organisé avec 13 participants.

Concernant l'organisation scientifique du congrès en elle-même, nous avons cette année reçu 37 soumissions :

- 23 articles pour la session « Recherche »
- 11 articles pour la session « Internationale »
- 3 propositions pour la session « Démos »

Les articles ont fait l'objet d'une évaluation par les membres du comité de programme (CP). A la suite de cette évaluation, les membres du conseil du comité de programme (CoP) ont cherché, avec les rapporteurs, à dégager une décision consensuelle qui a été présentée lors de la réunion de sélection des articles. Tout ce travail d'évaluation et d'échanges est un travail très important qui plus est parfois réalisé dans des délais relativement courts.

Suite à la réunion de sélection des articles, nous avons retenu pour cette année :

- 12 articles pour la session « Recherche » (10 articles longs et 2 articles courts)
- 12 articles pour la session « Internationale »
- 4 propositions pour la session « Démos »

*Ne soyez pas étonné(e) des chiffres ci-dessus. Certains articles ont en effet été retenus dans une session différente de celle de soumission.*

Sur la base des articles retenus, le programme du congrès est construit et organisé autour des thématiques principales suivantes :

- Méthodes
- Gouvernance & SI
- Ingénierie des Documents et des Connaissances
- Modélisation pour les SI
- SI & Données massives (Big Data)
- SI & Responsabilité Sociétale et Environnementale
- SI & Sciences des données
- SI & Sécurité

En plus de ces sessions thématiques, une session dédiée permet aux participants d'avoir, par leurs auteurs, une démo live de leurs outils, plateformes ou expérimentations.

Les participants au congrès INFORSID 2024 ont également pu avoir une présentation de la thèse de *Landy Andriamampianina*, prix de thèse 2024 attribué par l'association INFORSID et intitulée « ***Temporal Graphs: From Modelling to Analysis*** ».

Le programme du congrès intègre également, le premier jour, deux ateliers participatifs :

- « ***Systèmes d'Information Responsables*** » animé par *Rebecca Deneckere, Elena Kornyshova et Nathalie Valles-Parlangeau*

- « ***Impacts*** » animé par *Pierre-Emmanuel Arduin, Raphaëlle Bour et Cecile Favre*.

Nous avons en outre le plaisir d'accueillir deux conférencières invitées sur des sujets importants et d'actualité :

- « **Manipulation de Données Structurées et Interaction avec des Outils Externes grâce aux LLMs : L'Avenir de l'IA Générative dans les Systèmes d'Information ?** » présentée par *Laure Soulier*

- « **Large Scale Trustworthy Distributed Collaborative Systems: Challenges and Prospective Solutions** » présentée par *Claudia-Lavinia Ignat*.

Enfin, avant de clore cette préface, je me dois de remercier toutes les personnes qui ont permis que ce colloque ait lieu :

- Les membres du bureau de l'association INFORSID qui m'ont fait l'honneur de me confier la tâche de Président du Comité de Programme. Je les remercie vivement pour la confiance qu'ils m'ont témoignée et l'aide continue qu'ils m'ont apportée tout au long de ces derniers mois.
- Les porteuses et porteurs des ateliers qui participent, à leur échelle, à l'animation de la science. Je tiens à particulièrement souligner leur choix d'une démarche participative tout à fait adaptée aux thématiques traitées.
- L'ensemble des auteures et auteurs qui permettent de nous éclairer sur ce que sont aujourd'hui, de leur point de vue, les problématiques liées aux SI.
- L'ensemble des membres du comité de programme pour leur travail d'évaluation qui ont permis de produire des revues riches et de qualité.
- L'ensemble des membres du conseil du comité de programme pour leur animation des discussions et leur participation à la réunion de sélection des articles.
- *Laure Soulier* et *Claudia-Lavinia Ignat* pour avoir accepté de venir partager leur point de vue sur leur domaine d'expertise scientifique en lien avec les SI.
- *Mario Cortes-Cornax*, responsable du forum JCJC pour son implication dans l'organisation de ce forum. En effet, je crois fermement que l'une des forces des sociétés savantes nationales est ces « jeunes ».

**Mais que serait un congrès sans son « tiers-lieu » ?**

Je remercie très chaleureusement l'ensemble des membres du comité d'organisation au travers de son président *Khalid Benali* et de son vice-président *Sylvain Castagnos* pour l'organisation de ce congrès à Nancy. Je les remercie pour leur implication de tous les instants et pour le fait qu'ils aient su répondre présents lorsque cela le nécessitait. A travers eux, je tiens également à remercier l'ensemble des soutiens (institutionnels ou industriels) qui ont permis que le congrès ait lieu dans les meilleures conditions qu'il soit.

Enfin, au-delà de toutes les personnes citées, je remercie l'ensemble des participantes et participants à ce congrès et qui font vivre, par leur présence, la communauté INFORSID.

***Merci à toutes et à tous !***

*Max CHEVALIER*

*Président du comité de Programme*

# Membres du Comité d'Organisation (CO)

**Président :**

- Khalid BENALI, LORIA, Université de Lorraine

**Vice-Président :**

- Sylvain CASTAGNOS, LORIA, Université de Lorraine

**Membres :**

- Gérôme CANALS, LORIA, Université de Lorraine
- François CHAROY, LORIA, Université de Lorraine

**Communication :**

- Marie BARON, LORIA

**Budget :**

- Nathalie FRITZ, LORIA

**Service de Gestion :**

- Delphine HUBERT, LORIA

# Membres du Comité de Programme (CP)

Lylia ABROUK, LIB, Université de Bourgogne

Pascal ANDRÉ, LS2N, Nantes Université

Ladjel BELLATRECHE, LIAS, Université de Poitiers

Mireille BLAY-FORNARINO, I3S, Université Côte d'Azur

Sylvain CASTAGNOS, LORIA, Université de Lorraine

Arnaud CASTELLTORT, LIRMM, Polytech Montpellier

François CHAROY, LORIA, Université de Lorraine

Adrian-Gabriel CHIFU, LIS, Université d'Aix-Marseille

Camelia CONSTANTIN, LIP6, Sorbonne Université

Nadine CULLOT, LIB, Université de Bourgogne

Rebecca DENECKERE, CRI, Université Paris 1 Panthéon Sorbonne

Cédric DU MOUZA, CEDRIC, CNAM

Cyril FAUCHER, L3i, La Rochelle Université

Cécile FAVRE, ERIC, Université Lyon 2

Faiza GHOZZI, MIRACL, Institut supérieur d'informatique et de multimédia, SFAX

Gérald KEMBELLEC, CNAM

Sébastien LABORIE, LIUPPA, Université de Pau et des Pays de l'Adour

Nicolas LABROCHE, LIFAT, Université de Tours

Anne LAURENT, LIRMM, Université de Montpellier

Sabine LOUDCHER, ERIC, Université Lyon 2

Sofian MAABOUT, LaBRI, Université de Bordeaux

Maude MANOUVRIER, LAMSADE, Université Paris-Dauphine-PSL

Kathia MARÇAL DE OLIVEIRA, LAMIH - Université Polytechnique Hauts-de-France

Imen MEGDICHE, IRIT, ISIS, Institut National Universitaire Champollion (INUC)

Elsa NEGRE, LAMSADE, Université Paris-Dauphine-PSL

Noel NOVELLI, LIS, Université d'Aix-Marseille

Andre PENINOU, IRIT, Université Toulouse 2

Thomas POLACSEK, ONERA

Christian SALLABERRY, LIUPPA, Université de Pau et des Pays de l'Adour

Marinette SAVONNET, LIB, Université de Bourgogne

Jiefu SONG, IRIT, Université Toulouse Capitole

Marlène VILLANOVA, LIG, Université Grenoble Alpes

Cédric WEMMERT, ICube, Université de Strasbourg

---

## Membres du Conseil du Comité de Programme (CoP)

---

Khalid BENALI, LORIA, Université de Lorraine

Agnes FRONT, LIG, Université Grenoble Alpes

Régine LALEAU, LACL, Paris-Est Créteil

André MIRALLES, UMR Tetis, INRAE

Christophe PONSARD, CETIC, Université de Namur, Belgique

Jolita RALYTE, CUI, Université de Genève, Suisse

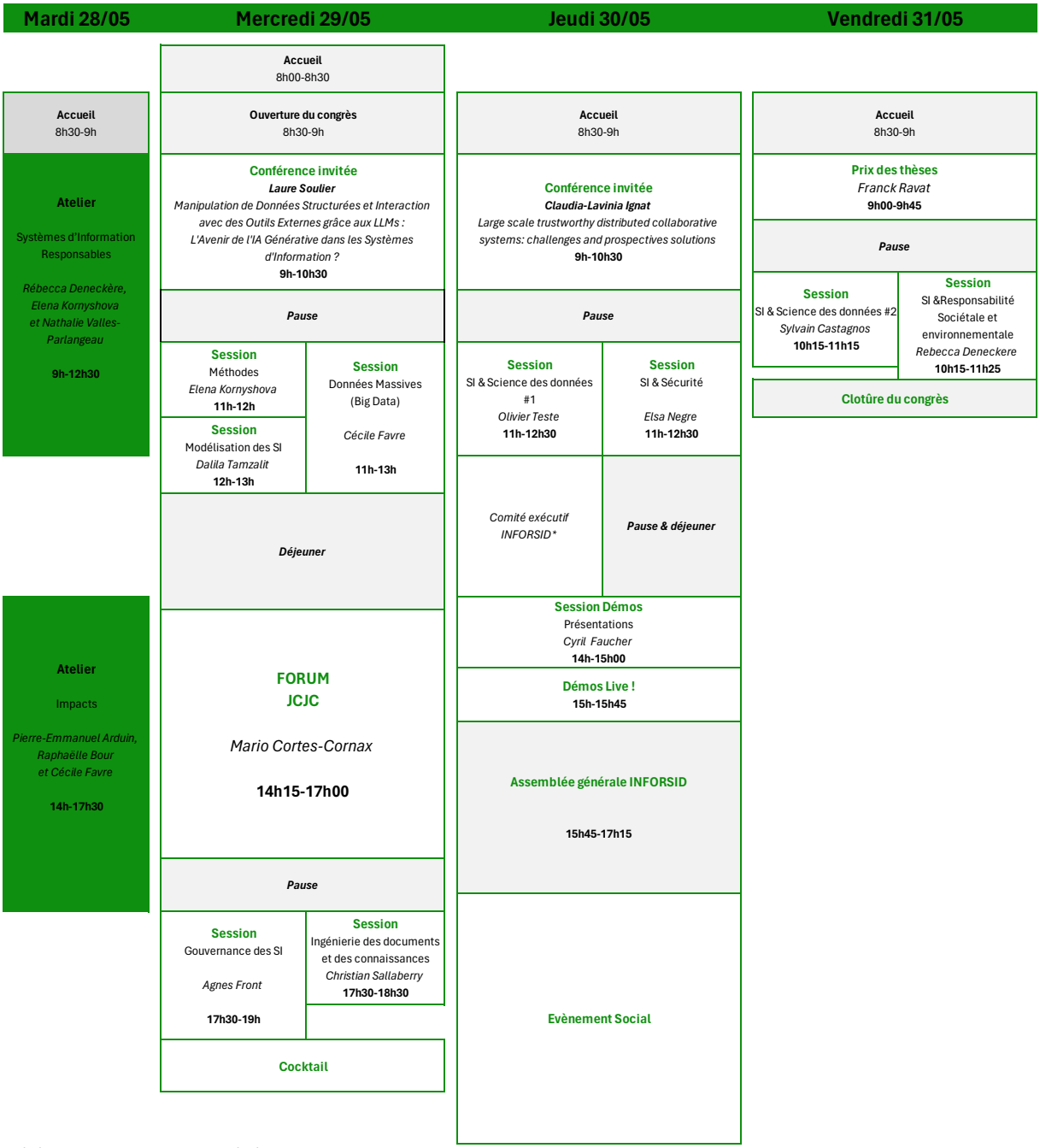
Philippe ROOSE, LIUPPA, Université de Pau et des Pays de l'Adour

Camille SALINESI, CRI, Université Paris 1 Panthéon-Sorbonne

Chantal SOULÉ-DUPUY, IRIT, Université Toulouse Capitole

Olivier TESTE, IRIT, Université Toulouse 2

# Programme synthétique INFORSID 2024




* dédié exclusivement aux membres du comité exécutif d'INFORSID

# Prix de thèse 2024 de l'association INFORSID

L'association INFORSID félicite *LANDY ANDRIAMAMPIANINA* pour sa thèse soutenue au cours de l'année 2023 et élue ***Prix de Thèse 2024***.

**Titre de la thèse** : Temporal Graphs: From Modelling to Analysis (résumé)

**Laboratoire** : Institut de Recherche en Informatique de Toulouse (UMR 5505)

**Directeurs de thèse** : Franck Ravat et Nathalie Valles-Parlangeau

**Date de soutenance** : 06/12/2023

**Mots-clefs** : Temporal Graphs, Modelling, Querying, Knowledge Discovery

# Table des matières des actes

## « Conférences invitées » - Animation Max Chevalier, Khalid Benali



## « Méthodes » - Animation Elena Kornyshova



## « Modélisation des SI » - Animation Dalila Tamzalit



## « SI & Données Massives (Big Data) » - Animation Cécile Favre



## « Gouvernance & SI » - Animation Agnès Front

***Légende***

Internat. **Papier International**

Long **Papier Long**

Court **Papier Court**

Démo **Démo**

# Manipulation de Données Structurées et Interaction avec des Outils Externes grâce aux LLMs : l'Avenir de l'IA Générative dans les Systèmes d'Information ?

**Dr. Laure Soulier**

*Sorbonne Université*
*laure.soulier@sorbonne-universite.fr*

## 1. Résumé

L'IA générative impacte de nombreux domaines d'application en entreprise. En premier lieu, et surtout grâce à ChatGPT, perçue comme un formidable outil pour interagir avec les utilisateurs, l'IA générative offre également de nombreuses perspectives de généralisation et d'adaptation.

Nous explorerons dans cette présentation les rouages des (grands) modèles de langue (LLMs en anglais), modèles fondamentaux de l'IA générative, leurs capacités à résoudre des tâches de plus en plus complexes, ainsi que leurs limites.

Nous aborderons ensuite un point de vue d'utilisation des LLMs pour les systèmes d'information autour de la valorisation des données structurées (bases de données, graphes de connaissances, ...) et l'exploitation d'outils externes/API pour la synthèse d'information.

## 2. Biographie

Laure Soulier[1] est maîtresse de conférences à Sorbonne Université au sein de l'équipe « Machine Learning and Information Access » de l'Institut des Systèmes Intelligents et de Robotique. Elle s'intéresse aux techniques d'apprentissage profond pour des tâches de traitement automatique du langage et de recherche d'information.

[1] https://pages.isir.upmc.fr/soulier/

Sa recherche s’articule autour de trois thématiques : la génération de résumés textuels à partir de données structurées ("data-to-text"), les moteurs de recherche conversationnels, et récemment l'exploitation des gros modèles de langue pour la robotique.

Elle co-publie des articles scientifiques dans les conférences et journaux internationaux de renommées en machine learning (NeurIPS, ICLR, ICML, AAAI) et en traitement automatique de la langue/recherche d'information (EMNLP, EACL, SIGIR, ECIR, CIKM).

Elle est/a été impliquée dans de nombreux projets européens ou nationaux, dont l'ANR JCJC SESAMS qu'elle coordonnait.

# Large scale trustworthy distributed collaborative systems: challenges and prospective solutions

**Dr. Claudia-Lavinia IGNAT**

*Inria Centre at Université de Lorraine*
*claudia.ignat@inria.fr*

## 1. Résumé

Computer-mediated collaboration is now part of both professional and personal spheres, facilitated by advancements in mobile and ubiquitous communication technologies and the widespread adoption of existing tools by users. The ubiquity of remote work has accentuated the reliance on computer-mediated collaborative tools across various domains such as work, education, and entertainment. This reliance is expected to grow further due to the ongoing digitalization of activities and the imperative to reduce travel in response to global warming.

While current computer-mediated tools effectively support small groups engaged in well-organized collaboration, they fall short when it comes to facilitating cooperation among large and diverse groups and organizations working on more extensive projects over extended durations.

Existing collaborative systems face several challenges including privacy concerns arising from the control of personal user data by major corporations, with users having limited influence over the utilization of their information. Performance and security issues are also notable, particularly when considering the scale of these collaborative systems.

This presentation outlines our vision for the development of trustworthy distributed collaborative systems, wherein communities of users can engage in collaborative endeavors securely and confidently without the necessity of a centralized authority. The discussion will primarily delve into anticipated solutions for issues related to replicated data consistency, security, and trust within the context of large-scale collaboration. Recognizing the pivotal role of the human factor in designing

trustworthy distributed collaborative systems, we emphasize the imperative of evaluating these systems through user studies.

## 2. Biographie

Claudia-Lavinia Ignat[1] is a tenured research scientist at Inria and head of the Coast team. She obtained a PhD in Computer Science from ETH Zurich, Switzerland and an habilitation (HDR) from Lorraine University. Her research domain is distributed collaborative systems that enable distributed group work using computer technologies. Designing such systems requires an expertise in distributed systems and computer-supported cooperative work. Besides theoretical and technical aspects of distributed systems, design of distributed collaborative systems must take into account the human factor to offer suitable solutions for users.

Her work is organized around three axes of research: collaborative data management referring to the design and evaluation of various approaches related to the management of distributed shared data including data replication and group awareness; security mechanisms for distributed collaborative systems without central authority; trustworthy collaboration referring to the evaluation of trust in collaborators.

She was general co-chair and PC member co-chair of ECSCW 2018 international conference. She is an associate editor of Journal of Computer Supported Cooperative Work and a regular PC member of the CHI, CSCW, ECSCW, GROUP, CollabTech and ICCP conferences. She was the coordinator of the USCOAST associated Inria team in collaboration with the Department of Psychology from Wright State University. She is the coordinator of Alvearium Inria challenge project on peer-to-peer cloud storage with hive enterprise and of the IPCEI DXP project on a federated and distributed data exchange platform with Amadeus. She also co-coordinates the PILOT project of PEPR eNSEMBLE.

[1] https://members.loria.fr/CIgnat/

# T-AGILE : Gestion de Projet Agile en Télétravail

**David Serruya, Rébecca Deneckère**

*Centre de Recherche en Informatique*
*Université Paris 1 Panthéon-Sorbonne, Paris, France*
*denecker@univ-paris1.fr*

*REFERENCE DE L'ARTICLE INTERNATIONAL Cet article est une synthèse de l'article : David Serruya, Rébecca Deneckère: T-AGILE: A Guide to Teleworking in Agile Project Management. KES 2023: 1997-2007.*

## 1. Introduction

La crise sanitaire du COVID-19 a contraint un grand nombre de personnes à opter pour le télétravail, alors qu'il était autrefois considéré comme un luxe. Pour certains, il est devenu une nécessité ou une alternative viable au travail sur site. Par le passé, les managers ne voyaient pas nécessairement le télétravail d'un bon œil mais la pandémie a permis au télétravail de devenir une alternative possible pour les travailleurs habitués à se rendre sur leur lieu de travail chaque jour. En 2020, lors du premier confinement en France, le télétravail a connu une croissance sans précédent, atteignant 41 % des employés, et il reste très courant, avec 38 % des employés travaillant à distance à la fin de 2021[1]. De plus, un nouveau modèle émerge, le mode hybride, où les employés alternent entre le travail en présentiel et à distance. 63 % des managers estiment que cette nouvelle forme de travail continuera à se développer, et 84 % souhaitent la déployer au sein de leur entreprise[2]. En raison de ces tendances, nous pouvons conclure que le télétravail ou ses variantes deviennent la norme dans de nombreuses entreprises.

Au cours des deux dernières décennies, les méthodes agiles de gestion de projet ont gagné en popularité auprès de nombreuses entreprises dans le domaine de la technologie de l'information. Le Manifeste Agile a été créé lorsque dix-sept experts en développement logiciel se sont réunis pour discuter des similitudes entre leurs différentes méthodes de développement (Beck *et al.,* 2001). Avec le télétravail

[1] https://newsroom.malakoffhumanis.com/actualites/malakoff-humanis-presente-les-resultats-de-son-barometre-teletravail-et-organisations-hybrides-2022-0686-63a59.html
2 https://www.oecd.org/coronavirus/policy-responses/le-teletravail-pendant-la-pandemie-de-covid-19-tendances-et-perspectives-e76db9dd/

devenant de plus en plus présent, il est important de remettre en question la faisabilité des organisations agiles où la communication et les interactions humaines sont cruciales. Nous nous interrogeons donc dans ce travail sur la question de savoir si la gestion de projet agile fait partie des pratiques facilitées ou, au contraire, contraintes par le travail à distance. Nos recherches nous ont amenés à compîler des entretiens avec des experts, les réponses à un sondage et la littérature existante sur le sujet. Nous proposons un guide destiné aux chefs de projet afin de vérifier si tout a été mis en place pour faciliter la gestion de projet agile en télétravail.

## 2. T-Agile, ou comment être agile en télétravail

Un guide contenant 9 bonnes pratiques a été proposé. Chacune indique l'objectif principal, les tâches à effectuer, le responsable principal, les personnes concernées, les outils possibles et la période du projet impactée. Ce guide a été validé avec un autre expert travaillant sur un projet en cours au sein d'une entreprise française.

Les bonnes pratiques proposées sont : (A) assurer une adaptation adéquate à la distanciation ; (B) apprendre à communiquer à distance ; (C) renouer avec des personnes en dehors de l'équipe (clients, partenaires commerciaux, autres équipes) ; (D) adapter les réunions au distanciel ; (E) améliorer l'expérience utilisateur des outils de suivi de projet ; (F) maintenir la gestion visuelle de l'équipe ; (G) mettre en place la programmation en binôme ; (H) adapter les rituels et cérémonies agiles ; (I) créer des moments informels.

## 3. Conclusion

Certains aspects du télétravail sont facilités par l'agilité et ses principes, en faisant une technique très adaptée au travail à distance. Les principes du Manifeste Agile et les valeurs agiles, telles que l'acceptation du changement, l'amélioration continue, et la création d'un environnement d'équipe sain et transparent, en sont d'excellents exemples. Du point de vue de l'agilité, le travail hybride est la solution la plus optimale car il conserve les avantages du télétravail tout en préservant les interactions et la communication, parfois difficiles à gérer à distance, en revenant au bureau physique de manière occasionnelle.

Nous avons proposé T-Agile, un guide pour l'Agilité en Télétravail, afin de garantir une bonne gestion de projet agile lors de l'introduction du télétravail dans une équipe, et nous l'avons validé avec un expert, obtenant de bons résultats. Face aux résultats obtenus, il est légitime de se demander si l'agilité en télétravail, en particulier le travail hybride, deviendra la norme dans la gestion de projets informatiques à long terme.

## Bibliographie

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... & Thomas, D. (2001). *The agile manifesto*.

# Une approche low-code pour la création, l'adaptation et l'exécution des méthodes

**Raquel Araújo de Oliveira[1], Mario Cortes-Cornax[2], Agnès Front[3]**

*Univ. Grenoble Alpes, CNRS, Grenoble INP[1], LIG*
*38000 Grenoble, France*
*1. raquel.oliveira@univ-grenoble-alpes.fr*
*2. mario.cortes-cornax@univ-grenoble-alpes.fr*
*3. agnes.front@univ-grenoble-alpes.fr*

*Cet article est une synthèse de l'article :*
*Oliveira, R.A., Cortes Cornax, M., Front, A.: Supporting Method Creation, Adaptation and Execution with a Low-code Approach. BPMDS/EMMSAD@CAiSE 2023: 184-198*

## 1. Introduction

L'ingénierie des méthodes est définie comme la discipline permettant de concevoir, construire et adapter des méthodes pour le développement de systèmes d'information (Brinkkemper, 1996). Cette discipline favorise l'adaptation des méthodes à des besoins, contextes ou projets particuliers. Pour faciliter la création de méthodes, une des approches consiste à proposer des frameworks de méthodes permettant de servir de modèles à partir desquels une nouvelle méthode peut être créée. Bien que de tels frameworks fournissent une aide considérable pour la création d'une méthode, peu de guidage est proposé sur l'utilisation de ces frameworks et ensuite sur l'utilisation de la méthode. En effet, une fois la méthode créée, une exécution efficace de la méthode nécessite le support d'outils pour exécuter chaque étape de la méthode et centraliser les résultats de chaque étape.

Dans cet article, nous présentons l'application d'une approche low-code pour *la création* de frameworks de méthodes, la création de méthodes à partir de ces frameworks et l'exécution de ces méthodes. Le paradigme low-code favorise le développement d'applications avec peu de codage pour livrer rapidement des applications (Richardson et al., 2014). Basé sur l'utilisation de composants pré-compilés, il s'adresse en priorité à des utilisateurs non spécialistes en IT.

[1] Institute of Engineering Univ. Grenoble Alpes

## 2. Présentation de l'approche low-code pour l'ingénierie de méthodes

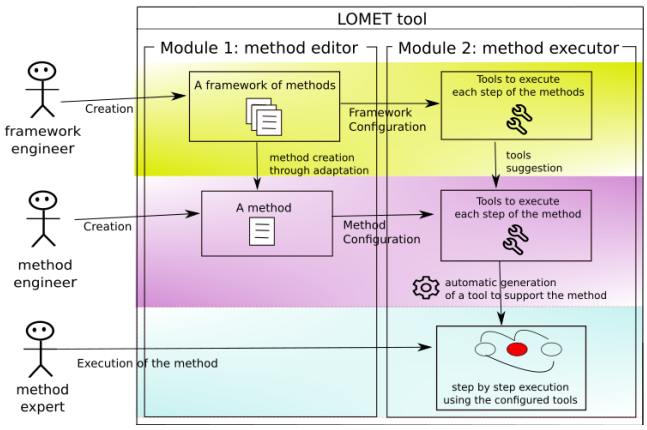



*Figure 1. L'approche low-code pour l'ingénierie de méthodes*

Premièrement, un *ingénieur de frameworks* crée un framework de méthodes (Figure 1), qui facilite la création ultérieure d'une méthode. Un tel framework définit un cadre générique de modèles de processus et un méta-modèle de produit, adaptables à de nombreux domaines d'application. L'ingénieur de frameworks peut également configurer le framework en paramétrant les outils les plus adaptés pour exécuter chaque étape des méthodes.

Ensuite, les *ingénieurs méthodes* créent une méthode en adaptant un framework de méthodes, ce qui signifie que le modèle de processus et le méta-modèle produit du framework peuvent être étendus pour créer une nouvelle méthode, en suivant le cadre prédéfini. Alternativement, l'ingénieur méthodes peut également créer une méthode à partir de zéro. La liste des outils configurés dans le framework est ensuite suggérée pour configurer la méthode nouvellement créée (dans le cas où cette dernière a été créée à partir d'un framework). L'ingénieur méthodes peut également proposer de nouveaux outils pour exécuter une étape d'une méthode.

Une fois la méthode créée et configurée, *l'expert méthode* peut mener l'exécution de la méthode étape par étape. Suivant le paradigme du low-code, un outil support pour exécuter la méthode est généré automatiquement et peut être utilisé par les experts méthodes afin de suivre la progression et les différents artefacts qui résultent de chaque étape. L'approche low-code est ainsi mise en œuvre pour l'expert-méthode qui bénéficie d'un outil d'exécution de la méthode, sans avoir lui-même à le développer.

### Bibliographie

Brinkkemper, S. (1996). *Method engineering: engineering of information systems development methods and tools*. Information and software technology 38(4), 275–280.

Richardson, C., Rymer, J.R., Mines, C., Cullen, A., Whittaker, D. (2014). *New development platforms emerge for customer-facing applications*. Forrester: Cambridge, MA, USA 15.

# La modélisation conceptuelle

## Passé, présent et futur

**J. Akoka[1], I. Comyn-Wattiau[2], N. Prat[2], V. C. Storey[3]**

*1. Laboratoire CEDRIC-CNAM*
*2. ESSEC Business School*
*3. Georgia State University, Atlanta, GA, USA*

*REFERENCE DE L'ARTICLE INTERNATIONAL. Cet article est une synthèse de l'article : Jacky Akoka, Isabelle Comyn-Wattiau, Nicolas Prat, Veda C. Storey: The journey of conceptual modeling: Paths from the past to present with trajectories for the future. ER (Companion) 2023.*

*MOTS-CLES : modélisation conceptuelle, analyse bibliographique de thèmes, analyse des chemins principaux, analyse des co-citations, analyse du couplage bibliographique.*

*KEYWORDS: conceptual modeling, bibliographic topic analysis, main path analysis, co-citation analysis, bibliographic coupling analysis.*

La modélisation conceptuelle, apparue depuis plus de cinq décennies, continue de connaître des développements importants. Pour comprendre comment ce domaine évolue, nous avons procédé à des analyses bibliométriques d'un corpus de 4652 articles de recherche couvrant la période 1976-2022.

Notre revue de littérature a établi qu'aucune étude antérieure ne couvre la période allant de 1976 à 2022 ni n'exploite les techniques de couplage bibliographique ni d'analyse des chemins principaux pour la modélisation conceptuelle.

Nous avons interrogé la base de données bibliographiques Scopus, à l'aide des mots-clés « conceptual mode(l)ling » et « entity-relationship model ». Nous avons sélectionné aussi tous les articles publiés lors de la conférence internationale sur la modélisation conceptuelle (ER) et les actes des ateliers associés.

L'analyse des co-citations identifie une structure intellectuelle du domaine en quatre composantes. La première débute avec l'introduction du modèle entité-association (Chen 1976), qui fournit un ensemble de concepts permettant d'abstraire le monde réel. La deuxième se focalise sur la qualité de la modélisation conceptuelle. Elle regroupe des contributions en matière de grammaires et de règles de guidage pour représenter le monde réel, ainsi que des méthodes d'évaluation de

ces représentations. La troisième composante rassemble des articles qui relèvent de la recherche en science de conception ainsi qu'à l'ingénierie des exigences. Enfin, la quatrième, plus récente, combine les ontologies et la modélisation conceptuelle.

Une analyse de couplage bibliographique sur la période 2017-2022 produit dix classes caractérisant les principaux thèmes de la recherche récente en matière de modélisation conceptuelle. La première classe décrit des approches de modélisation conceptuelle spécifiques à un domaine, permettant l'accumulation d'un ensemble de connaissances pour celui-ci, par exemple les villes intelligentes. La seconde classe se concentre sur les ontologies et leur projection sur des domaines d'application. La troisième classe est centrée sur l'ingénierie des exigences. La quatrième est dédiée à la modélisation de processus. La cinquième rassemble des contributions sur les bases de données, principalement le monde NoSQL. La classe 6 regroupe la modélisation conceptuelle appliquée à différents domaines, avec ou sans ontologies. La classe 7 traite du domaine de la santé et de la biologie. Les classes 8, 8 et 10 traitent respectivement de la gestion des connaissances, la modélisation conceptuelle multi-niveaux et les approches temporelles et économiques.

L'analyse des chemins principaux du réseau de citations révèle trois dynamiques successives dans l'évolution des articles influents, c'est-à-dire très cités et liés dans une chaîne de citations. La première dynamique, de "raffinement du modèle ER » s'étend de 1976 à 2000 avec l'article (Parsons et Wand, 2000) proposant une modélisation en classes et instances. La deuxième décrit des contributions visant "l'amélioration de la modélisation conceptuelle avec les ontologies ». La troisième dynamique, "modélisation conceptuelle pour le monde numérique", émerge en 2017, mettant l'accent sur les concepts plus récents de crowdsourcing, de plateformes ouvertes ou encore la combinaison des réalités physiques et numériques.

Cette étude bibliométrique décrit une approche empirique, dont la particularité est de se concentrer sur la catégorisation des recherches antérieures pour aider à comprendre la structure et la dynamique d'un domaine et ses évolutions possibles. Les résultats peuvent être imparfaits en raison de problèmes de qualité de données. Les techniques bibliométriques se concentrent sur les relations de citation et de cooccurrence entre les documents, ignorant la sémantique de ces relations. Des recherches futures permettront d'affiner les périodes identifiées.

## Bibliographie

Akoka J., Comyn-Wattiau I., Prat N., Storey V.C. (2023). The journey of conceptual modeling: Paths from the past to present with trajectories for the future. ER2023: Companion Proceedings of the 42nd International Conference on Conceptual Modeling: ER Forum, Lisbon, Portugal.

Chen, P.P.-S., The entity-relationship model—toward a unified view of data. ACM transactions on database systems (TODS), 1976. 1(1): p. 9-36.

Parsons J., Y. Wand (2000). Emancipating instances from the tyranny of classes in information modeling. ACM Transactions on Database Systems (TODS), 25(2): p. 228-268.

# Modélisation pour l'analyse et la conception dans les écologies d'artefacts réglementées (MADRAE)

## Analyse d'un cas à propos de pratiques coopératives en télémédecine

**Clément Cormi [1], Khuloud Abou Amsha[2], Matthieu Tixier[2], Myriam Lewkowicz[2]**

*1. Chaire Innovation du Bloc Opératoire Augmenté (BOPA), AP-HP, Institut Mines-Telecom, Université Paris Saclay, Villejuif, France. clement.cormi-ext@aphp.fr*
*2. Université de Technologies de Troyes, LIST3N. {prénom.nom}@utt.fr*

*REFERENCE DE L'ARTICLE INTERNATIONAL*
*Clément Cormi, Khuloud Abou Amsha, Matthieu Tixier, Myriam Lewkowicz: Modeling for Analysis and Design in Regulated Artifacts Ecologies (MADRAE): a Case for Cooperative Practices in Telemedicine. Proceedings of 21st European Conference on Computer-Supported Cooperative Work. (2023) DOI: 10.48340/ecscw2023_ep03.*

### 1. Introduction

La création de technologies efficaces pour le travail coopératif nécessite une compréhension nuancée des interactions sociales (Randall & al., 2007). Toutefois, malgré une analyse approfondie des dynamiques de coopération, l'impact des technologies nouvellement introduites reste souvent limité, avec des difficultés notables en matière d'adoption et d'appropriation par les utilisateurs. Cette situation persiste malgré l'existence d'analyses détaillées des pratiques de coopération soutenues par les outils numériques (Lewkowicz & Liron, 2019).

Dans le contexte du génie logiciel, la modélisation sociotechnique émerge comme une approche pertinente, associant les dimensions sociales, techniques, et organisationnelles du travail. Les recherches en CSCW (Computer-Supported Cooperative Work) ont proposé des méthodes de modélisation qui embrassent cette complexité pour la conception d'outils numériques alignés sur les besoins des utilisateurs (Divitini et al., 1996; Herrmann et al., 2004). Malgré leur potentiel, ces

approches n'ont pas été largement adoptées en génie logicielle et dans l'industrie, contrairement à UML.

Cet article propose une approche qui vise à rapprocher des travaux théoriques sur le travail coopératif soutenus par le numérique et les pratiques de conception en génie logiciel dans une perspective sociotechnique. En enrichissant les modèles de conception à l'appui de résultats de recherches de terrain en CSCW, nous cherchons à améliorer la prise en compte des pratiques de travail dans l'analyse et la conception logicielle. Deux concepts nous intéressent en particulier, celui d'écologies d'artefacts (Bødker et al., 2016; Larsen-Ledet et al., 2020) et celui de travail des données ou *data work* (Bossen et al., 2019).

Nous présentons MADRAE (Modeling for Analysis and Design in Regulated Artifacts Ecologies - https://github.com/Clement-Cormi/MADRAE), une extension du diagramme de composants UML (Cormi et al, 2023). Nous mobilisons une étude de cas en télémédecine pour montrer comment intégrer des aspects essentiels comme le travail des données et les écologies d'artefacts dans l'analyse, la conception et le déploiement de logiciels. Les retours d'une évaluation experte soulignent les intérêts et limites actuelles de la démarche MADRAE.

## Bibliographie

Bødker, S., Korsgaard, H., & Saad-Sulonen, J. (2016). A Farmer, a Place and at least 20 Members- The Development of Artifact Ecologies in Volunteer-based Communities. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16, 1140–1154

Bossen, C., Pine, K. H., Cabitza, F., Ellingsen, G., & Piras, E. M. (2019). Data work in healthcare: An Introduction. Health Informatics Journal, 25(3), 465-474.

Cormi, Clement; Abou-Amsha, Khuloud; Tixier, Matthieu; Lewkowicz, Myriam (2023): Modeling for Analysis and Design in Regulated Artifacts Ecologies (MADRAE): a Case for Cooperative Practices in Telemedicine. Proceedings of 21st European Conference on Computer-Supported Cooperative Work.

Divitini, M., Simone, C., & Schmidt, K. (1996). ABACO: Coordination mechanisms in a multiagent perspective. In COOP'96. Second International Conference on the Design of Cooperative Systems 103–122.

Herrmann, T., Hoffmann, M., Kunau, G., & Loser, K.-U. (2004). A modelling method for the development of groupware applications as socio-technical systems. Behaviour & Information Technology, 23(2), 119–135.

Larsen-Ledet, I., Korsgaard, H., & Bødker, S. (2020). Collaborative Writing Across Multiple Artifact Ecologies. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–14.

Lewkowicz, M., & Liron, R. (2019). The Missing "Turn to Practice" in the Digital Transformation of Industry. Computer Supported Cooperative Work (CSCW), 28(3–4), 655–683.

Randall, D., Harper, R., & Rouncefield, M. (2007). Fieldwork for design: theory and practice.

# EX-LAD : Un Tableau de Bord Explicable pour l'Analyse de lApprentissage dans l'Enseignement Supérieur

**Tesnim KHELIFI[1], Nourhène BEN RABAH[1], Bénédicte LE GRAND[1]**

*1. Centre de Recherche en Informatique, Université Paris 1 Panthéon Sorbonne*
*90 rue de Tolbiac, 75013 Paris, France*
*{Tesnim.Khelifi|Nourhene.Ben-Rabah|Benedicte.Le-Grand}@univ-paris1.fr.*

*RESUME « Cet article est une synthèse de l'article : Tesnim Khelifi,Nourhène Ben Rabah., Bénédicte Le Grand , Ibtissem Daoudi : EX-LAD : Explainable Learning Analytics Dashboard in Higher Education Proceedings of 36th International Conference on Computer Applications in Industry and Engineering, vol.97, pp. 38-51 »*

Depuis la pandémie de COVID-19, l'enseignement à distance s'est développé au sein des établissements d'enseignement supérieur (Schneider et al. 2021), souvent par le biais de LMS (Learning Management System) tels que Moodle ou BlackBoard Learn. Ces systèmes, outre l'accès à des ressources pédagogiques et des possibilités d'évaluation, permettent d'enregistrer les interactions entre les étudiants et la plateforme. Ces traces numériques peuvent être analysées et représentées sous forme de tableaux de bord, afin de faciliter le suivi des étudiants et éviter les situations d'échec ou d'abandon. Néanmoins, la plupart d'entre eux se concentrent sur les performances des étudiants au détriment d'autres indicateurs importants tels que l'engagement comportemental et émotionnel (Shohag et al. 2022), ce qui retarde la détection de leurs difficultés. De plus, les graphiques proposés sont parfois difficiles à interpréter, ce qui peut aboutir à des conclusions erronées ou à des actions de remédiation mal ciblées.

Nous proposons le tableau de bord EXplainable Learning Analytics Dashboard (EX-LAD), visant à aider les étudiants à s'auto-évaluer et à améliorer leur parcours d'apprentissage, à travers les indicateurs de performance, d'engagement et de persévérance tout en fournissant aux enseignants les outils dont ils ont besoin pour suivre leurs progrès et détecter ceux qui sont en difficulté et risquent d'échouer, afin d'intervenir au bon moment. Nous avons veillé à ce que le tableau de bord puisse être compris par toutes les personnes concernées, afin d'en maximiser l'efficacité.

Nous avons mené une étude de cas avec des données réelles collectées en 2021-2022 auprès de 128 étudiants de l'école d'informatique ESIEE-IT, lors d'un cours

Python enseigné de manière hybride. Les données ont été anonymisées conformément aux principes éthiques du RGPD.

Notre proposition de tableau de bord intègre diverses visualisations. En utilisant différents formats tels que des diagrammes à barres pour visualiser par exemple les moyennes et rangs des étudiants et des diagrammes de dispersion pour comprendre la dispersion des profils des étudiants, nous visons à fournir une vue d'ensemble complète et à faciliter les comparaisons (Figure 1). Les graphiques sont choisis en fonction de leur clarté et de leur accessibilité pour les utilisateurs, et mettent en œuvre une description textuelle et un code couleur pour en faciliter l'interprétation.

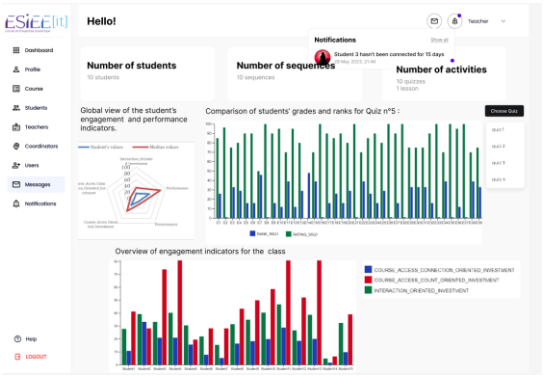



*Figure 1. Interface de l'enseignant*

Ce tableau de bord permet également de classifier les étudiants en quatre profils en fonction de leur score de performance et d'engagement. Nous identifions quatre profils d'étudiants : E+P+ pour ceux qui sont à la fois engagés et performants, E+P- pour les étudiants engagés mais non performants, E-P+ pour ceux qui sont performants mais non engagés, et enfin E-P- pour ceux qui sont à la fois non engagés et non performants. Nous avons proposé des pistes de remédiation adaptées à chacun de ces profils, afin de faciliter la tâche d'accompagnement pour l'enseignant.

Nos travaux futurs mettront en œuvre des techniques de Machine Learning pour prédire l'évolution des performances des étudiants et anticiper encore mieux leurs difficultés.

## Références

S. L. Schneider et M. L. Council, Distance learning in the era of COVID-19, (2021) *Arch Dermatol Res*, vol. 313, no 5, pp. 389-390

S. Shohag et M. Bakaul, A Machine Learning Approach to Detect Student Dropout at University, (2022), *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 10, p. 3101-3107

# Une approche pour l'extension à la demande de cubes multidimensionnels dans un contexte multi-modèles : Application à l'agroécologie basée sur l'internet des objets

**Sandro Bimonte[1], Fagnine Alassane Coulibaly[1], Stefano Rizzi[2]**

*1. TSCF-INRAE, Université de Clermont Auvergne*
*9 avenue Blaise Pascal, 63170 Aubière, France*
*prenom1-prenom2.nom@inrae.fr*

*2. DISI, Université de Bologne*
*2 Viale del Risorgimento, 40136 Bologna, Italie*
*prenom.nom@unibo.it*

*Cet article est une synthèse de l'article : Sandro Bimonte, Fagnine Alassane Coulibaly, Stefano Rizzi. An approach to on-demand extension of multidimensional cubes in multi-model settings: Application to IoT-based agro-ecology, Data & Knowledge Engineering, Volume 150, 2024, 102267, ISSN 0169-023X, https://doi.org/10.1016/j.datak.2023.102267*

La gestion, l'intégration et l'analyse des données semi et non structurées représentent des défis majeurs dans les écosystèmes de données. Ces systèmes doivent également être capables d'évoluer de manière transparente et efficace. Cette problématique est d'autant plus pertinente pour les analyses de type OLAP, qui exploitent des cubes multidimensionnels au sein des entrepôts de données (data warehouse, DW). Les DW stockent en effet une large variété de données, en s'appuyant sur divers modèles de données. Si la question de la gestion de cette diversité de modèles, via l'utilisation d'entrepôts de données multi-modèles (Bimonte et al., 2022), a été partiellement abordée, les enjeux liés à l'extensibilité des systèmes restent peu explorés.

L'extensibilité se réfère à la capacité d'intégrer de nouveaux éléments multidimensionnels dans les entrepôts de données multi-modèles (multi-models data warehouse, MMDW), qui n'étaient pas prévus ou identifiés lors de leur conception initiale, afin de les rendre disponibles pour des analyses OLAP ultérieures. Cela représente un avantage clé des MMDW par rapport aux DW purement relationnel. Cette caractéristique est particulièrement importante dans le contexte de l'émergence de nouvelles technologies de collecte de données issue de l'Internet des Objets (IoT).

Dans (Coulibaly et al., 2023), nous avons exploré une méthode pour implémenter l'extensibilité des cubes multidimensionnels, en adoptant une approche de définition de schéma lors de la lecture des données (Schema-on-Read). Cette technique se distingue de l'approche conventionnelle de schéma à l'écriture (Schema-on-Write), car elle permet de spécifier le schéma des données au moment où elles sont lues, au lieu de le faire lors de leur enregistrement. Une architecture a été proposée pour mettre en œuvre cette approche.

Cet article s'inscrit dans la continuité du précédent et présente « xCube », une méthode innovante favorisant l'extension dynamique des cubes multidimensionnels en réponse aux nouveaux besoins d'analyse et d'intégration de nouvelles données par les utilisateurs. « xCube » offre aux utilisateurs la possibilité de sélectionner un composant multidimensionnel (tel qu'un fait ou un niveau hiérarchique) à étendre avec des données additionnelles. Ces données additionnelles peuvent provenir soit d'un lac de données ou soit être déjà incluses dans le MMDW. Le processus d'extension du schéma multidimensionnel prend en compte les dépendances fonctionnelles entre ces nouvelles données et l'élément à étendre. En effet, ce processus d'extension utilise trois algorithmes spécifiques au type d'extension souhaité qu'il s'agisse de l'ajout de nouvelles mesures, dimensions ou niveaux hiérarchiques. Ces trois algorithmes s'appuient sur un quatrième algorithme qui renvoie l'ensemble des sommets candidats pour étendre l'élément du MMDW sélectionné. Cet algorithme recherche les dépendances fonctionnelles entre l'élément à étendre et les données additionnelles en parcourant les schémas relationnels, de documents et de graphes ainsi que les références inter-modèles. L'exécution de tous ces algorithmes permet d'obtenir un nouveau schéma multidimensionnel étendu. Ce schéma étendu est ensuite rendu accessible aux utilisateurs pour l'analyse OLAP.

Par ailleurs, nous avons proposé une implémentation de preuve de concept pour xCube dans le contexte de la surveillance épidémiologique de la Flavescence Dorée dans les vignobles de la région de Bordeaux. Nous avons utilisé AgensGraph pour le stockage de données multi-modèles et Mondrian comme serveur OLAP. Nous avons montré que l'utilisation des balises « View » et « MeasureExpression » de Mondrian permet d'utiliser comme dimensions, niveaux et mesures des données de type document et graphe stockées dans AgensGraph en encapsulant les requêtes de documents et de graphes dans des requêtes SQL. De cette manière, la complexité du stockage multi-modèle est transparente pour le serveur OLAP et, surtout, pour le client OLAP.

## Bibliographie

Bimonte, S., Gallinucci, E., Marcel, P., Rizzi, S. (2022). Data variety, come as you are in multi-model data warehouses. Information Systems 104, 101734. https://doi.org/10.1016/j.is.2021.101734

Coulibaly, F.A., Bimonte, S., Rizzi, S., Malembic-Maher, S., Fabre, F. (2023). Towards a Multi-Model Approach to Support User-Driven Extensibility in Data Warehouses: Agro-ecology Case Study. EDBT/ICDT Workshops 2023.

# Interrogation de Polystores hétérogènes multi-modèles à partir de Modèles Unifiés

**Léa EL AHDAB[1], André PENINOU[2], Olivier TESTE[2], Imen MEGDICHE[3]**

*1. Université de Toulouse UT3, IRIT, Toulouse, 118 Route de Narbonne, F-31062 Toulouse, France, lea.el-ahdab@irit.fr*
*2. Université de Toulouse UT2J, IRIT, Toulouse, 118 Route de Narbonne, F-31062 Toulouse, France, andre.peninou@irit.fr, olivier.teste@irit.fr*
*3. INU JF Champollion, IRIT, Avenue Georges Pompidou, F-81104 Castres, France, imen.megdiche@irit.fr*

*Cet article est une synthèse de l'article : El Ahdab, L., Teste, O., Megdiche, I., & Péninou, A. (2023, August). Unified views for querying heterogeneous multi-model polystores. In International Conference on Big Data Analytics and Knowledge Discovery, DAWAK 2023 (pp. 319-324). Cham : Springer Nature Switzerland.*

La problématique de cet article porte sur l'interrogation transparente de données distribuées de manière verticale dans des polystores. Systèmes de stockage flexibles, ils associent plusieurs bases de données SQL et NoSQL. Plusieurs solutions se développent en considérant la distribution verticale qui consiste à stocker des classes d'entités dans plusieurs bases de données différentes qui peuvent être de différents modèles. Certaines solutions se concentrent sur la représentation unifiée du polystore et proposent la migration des données vers un modèle unifié (cf. Barret *et al*., 2022). D'autres travaillent sur les opérateurs et utilisent des fonctions externes aux SGBD pour réaliser des jointures entre les systèmes (cf. Kolev *et al.*, 2016). L'hétérogénéité structurelle est peu considérée dans les systèmes NoSQL des polystores, de même que la fragmentation verticale des données qui correspond au stockage d'une classe d'entité dans différentes bases de données. La notion de clé de fragmentation est utilisée : c'est un attribut clé primaire présent dans l'ensemble des fragments identifiés et permettant de reconstruire l'entité fragmentée.

Cet article propose un Framework d'interrogation des polystores composés de bases de données Relationnelles (R) et Document (D). Il propose une représentation unifiée du polystore masquant l'hétérogénéité des modèles de données, la distribution et la fragmentation des données. Il se compose d'une phase de construction qui met en place les outils nécessaires à l'interrogation du polystore et d'une phase d'exploitation qui contient les étapes liées à la réécriture de la requête utilisatrice. *Phase de construction*. Elle se base sur des modèles unifiés par système (R ou D) correspondant à la traduction du modèle E/R du polystore dans ce système.

Chaque classe d'entité correspond à une table (R) ou à une collection (D). Les relations, selon leur cardinalité, peuvent correspondre à une table, à un attribut dans une table/collection (clé étrangère) ou à des attributs imbriqués dans les collections. Le polystore est alors représenté soit sous la forme de tables (R), soit de collections (D). Un dictionnaire de mapping fait le lien entre la position des attributs dans les modèles unifiés et leur position réelle dans le polystore. Un attribut dans un modèle unifié est associé à une table/collection. Dans le polystore, ce même attribut est associé à une base de données et à une table/collection. *Phase d'exploitation*. L'utilisateur écrit une requête $Q$ en SQL (opérateurs : $\sigma$, $\pi$, $\bowtie$) sur le modèle unifié relationnel ou en MongoDB (opérateurs : $match, $project) sur le modèle unifié document. $Q$ est analysée algébriquement et le dictionnaire de mapping permet de réécrire $Q$ avec la position réelle des attributs ($Q_{new}$). Des opérations sont ajoutées (jointure relationnelle ou lookup et unwind en document) dans le cas de la reconstruction d'une entité fragmentée ou dans le cas de changement de systèmes (transfert et transformation). Le résultat de l'exécution de $Q_{new}$ est renvoyé à l'utilisateur dans le format du modèle interrogé.

Pour les expérimentations, nous avons utilisé les données du dataset Unibench où les classes d'entités sont distribuées verticalement et certaines sont fragmentées. Plusieurs types de requête sont considérés selon les opérateurs ($\sigma$, $\pi$, $\sigma+\pi$) et le nombre de tables (mono-table, multi-tables). L'opérateur de jointure est associé aux requêtes multi-tables. Notre évaluation compare le temps de réécriture des requêtes sur les modèles unifiés (relationnel et document) en fonction de plusieurs distributions. Le temps de réécriture moyen des requêtes est d'environ 0.0041 secondes et leur temps d'exécution moyen est d'environ 10 secondes. Les modèles unifiés permettent de cacher la complexité du polystore en le représentant dans un format mono-store. L'utilisateur choisit donc le langage d'écriture de sa requête. La réécriture des requêtes est peu coûteuse par rapport au coût d'exécution qui dépend de la distribution et fragmentation des données. Des structures hétérogènes peuvent apparaître dans les systèmes de stockage document et notre système les prendra en compte dans la réécriture de $Q$ à partir de travaux existants.

*Remerciements :*

*Ces travaux ont été menés avec le soutien du Gouvernement Français dans le cadre du programme Territoire d'Innovation, une action du Grand Plan d'Investissement adossé à France 2030, de Toulouse Métropole et du GIS neOCampus.*

## Bibliographie

Barret N., Manolescu I., Upadhyay P. (2022). Abstra: Toward Generic Abstractions for Data of Any Model, *31st ACM International Conference on Information & Knowledge Management 2022,* Atlanta

Kolev B., Valduriez P., Bondiombouy C., et al. (2016). CloudMdsQL: querying heterogeneous cloud data stores with a common language. *Distributed and parallel databases* vol. 34, p. 463-503.

# Prévenir les erreurs techniques des analyses dans les data lakes avec la théorie des types

**Alexis Guyot, Éric Leclercq, Annabelle Gillet, Nadine Cullot**

*Laboratoire d'Informatique de Bourgogne (LIB), Université de Bourgogne*
*9 avenue Alain Savary, F-21078 Dijon CEDEX, France*
*{prenom}.{nom}@u-bourgogne.fr*

*REFERENCE DE L'ARTICLE INTERNATIONAL Cet article est une synthèse de l'article :*
*Alexis Guyot, Éric Leclercq, Annabelle Gillet, Nadine Cullot:*
*Preventing Technical Errors in Data Lake Analyses with Type Theory. DaWaK 2023: 18-24.*

Les *data lakes* sont des plateformes dédiées à l'analyse des données massives (Hai *et al.*, 2023). Ils fournissent un ensemble d'opérateurs pour explorer, transformer, enrichir et analyser les données à la demande. Ces opérateurs proviennent de différents outils spécialisés. Par exemple, des opérateurs de transformation peuvent provenir de SparkSQL ou de Pandas, des opérateurs d'analyse de TensorFlow ou de NetworkX.

Les utilisateurs des *data lakes* mettent en œuvre leurs analyses en composant ces différents opérateurs. On qualifie de telles compositions de *workflows* d'analyse. Les *workflows* d'analyse dans les *data lakes* sont généralement hybrides. Les opérateurs composés reposent sur différents cadres théoriques comme l'algèbre relationnelle, l'algèbre linéaire ou la théorie des graphes. Ils implémentent différents paradigmes de programmation comme l'orienté objet ou le fonctionnel. Ils peuvent s'exécuter dans différents environnements d'exécution comme un CPU, un GPU ou un DPU.

L'hybridité des *workflows* d'analyse est source d'erreurs techniques pouvant nuire à leur exécution. En effet, de nombreuses transformations intermédiaires sont nécessaires pour composer certains opérateurs : filtrage, jointure, changement de modèle, etc. Ces transformations s'appliquent sur différents niveaux d'abstraction, notamment sur les données elles-mêmes, sur leur schéma et sur leur modèle. Elles peuvent introduire des incohérences sur chacun de ces niveaux. Par exemple, une transformation peut retirer un attribut nécessaire à l'exécution d'un opérateur suivant du *workflow*. Une autre peut introduire un attribut dont le type n'est pas supporté par le modèle de données utilisé par un opérateur suivant.

Notre article présente une approche pour prévenir les erreurs techniques dans les *workflows* d'analyse des *data lakes* (Guyot *et al.,* 2023). Nous nous intéressons particulièrement aux erreurs techniques causées par des incohérences au niveau du schéma et du modèle. Notre approche transforme les erreurs techniques en erreurs

de types, de sorte à pouvoir les prévenir dès la compilation. Contrairement aux approches existantes, elle permet une représentation unifiée de différents modèles, de différents niveaux d'abstraction et des liens qui les unissent.

Pour cela, nous proposons un cadre formel basé sur la théorie des types (Martin-Löf et Sambin, 1984). Ce cadre définit un ensemble de types pour représenter les schémas et les modèles des entrées et sorties des opérateurs des *workflows*. La théorie des types est un formalisme constructif permettant la définition de types à partir de règles de construction. Elle fournit un ensemble de types primitifs : entiers (*Integer*), chaînes de caractères (*String*), etc. Elle fournit également un ensemble de constructeurs permettant la définition de types composites, notamment : des produits de types, notés *T1* x *T2* ; des types génériques, notés *T1[T2]* ; ou des types dépendants, notés $\Pi_{(x:T2)}T1(x)$. Un produit de types est un type dont les valeurs sont des paires de valeurs d'autres types, comme le type *Integer* x *String* de la valeur *(1, abc)*. Un type générique est un type paramétré par un autre type, comme le type *Array[Integer]* de la valeur *[1,2]*. Un type dépendant est un type paramétré par une valeur d'un autre type, comme le type $\Pi_{(3:Integer)}Array(3)$ de la valeur *[a,1,true]*.

Dans notre cadre formel, les schémas des données sont représentés par des produits de types dépendants-génériques, et les modèles par des types génériques. Ainsi, une relation Personne composée de deux attributs Nom et Age de types *String* et *Integer* peut être associée au type $Relation[\Pi_{(Nom:String)}Attribut[String](Nom)$ x $\Pi_{(Age:String)}Attribut[Integer](Age)]$. Les valeurs de ce type sont des tuples de chaînes de caractères et d'entiers comme *(guyot, 24)*. Les liens entre les deux niveaux d'abstraction sont exprimés au travers de règles de construction. Notre cadre — combiné avec les mécanismes de preuve de la théorie des types — permet de formellement prouver l'absence d'erreur dans une composition d'opérateurs.

Nous avons réalisé une implémentation des différentes constructions de notre cadre formel en Scala[1]. Celle-ci utilise le compilateur pour techniquement vérifier l'absence d'erreur. Elle repose sur les mécanismes de typage avancés du langage — notamment les implicites qui permettent de modifier le comportement du compilateur — et sur la bibliothèque Shapeless[2]. L'implémentation peut être utilisée pour spécifier des langages de domaine (*Domain Specific Language* ou DSL) permettant la définition de *workflows* d'analyse sûrs dans les *data lakes*.

**Bibliographie**

Guyot A., Leclercq E., Gillet A., Cullot N. (2023). Preventing Technical Errors in Data Lake Analyses with Type Theory, *International Conference on Big Data Analytics and Knowledge Discovery, DaWaK 2023 (pp. 18-24),* Springer, Penang, Malaysia.

Hai R., Koutras C., Quix C., Jarke M. (2023). Data Lakes : A Survey of Functions and Systems, *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, n°12, p. 12571-12590.

Martin-Löf P., Sambin G. (1984). *Intuitionistic type theory*, Bibliopolis, Naples.

[1] https://github.com/AlexisGuyot/type_safe_compo
[2] https://github.com/milessabin/shapeless

# La gouvernance des données en contexte universitaire : proposition d'un modèle de maturité

**Ugo Verdi[1,2], Nathalie Pinède[1] et Guy Melançon[2]**

*1. Laboratoire MICA, Université Bordeaux-Montaigne*
*Bâtiment MSH de Bordeaux, 10 Esplanade des Antilles, 33607 Pessac cedex*
*ugo.verdi@etu.u-bordeaux-montaigne.fr et nathalie.pinede@u-bordeaux-montaigne.fr*

*2. Laboratoire LaBRI, Université de Bordeaux*
*Domaine universitaire, 351 cours de la Libération, 33405 Talence*
*guy.melancon@u-bordeaux.fr et ugo.verdi@u-bordeaux.fr*

*RÉSUMÉ. Dans le contexte d'une production très rapide des données, la question de leur gouvernance traverse l'ensemble des organisations. Si cette problématique a été principalement étudiée dans un contexte entrepreneurial où une gestion optimale des données sert en priorité des objectifs économiques, elle a moins fait l'objet d'une étude dans un cadre universitaire. Cet article propose ici un modèle de maturité pour permettre aux universités de disposer d'un outil concourant d'une part à identifier et analyser les mécanismes de la gouvernance des données et, d'autre part, à évaluer le niveau d'implication des acteurs et de l'organisation. Bien qu'adapté à un contexte local particulier, à savoir les projets ACT de l'université de Bordeaux, il a pour vocation d'être adaptable à d'autres contextes universitaires.*

*ABSTRACT. As data production continues to surge, the challenge of data governance becomes increasingly pertinent across all organizations. Although extensively explored within business sectors, primarily for its economic benefits in data management, this topic has garnered less attention within university settings. This study introduces a maturity model specifically designed for universities. It aims to identify and analyze data governance mechanisms and assess stakeholder engagement and organizational involvement. Developed within the context of ACT projects at the University of Bordeaux, this model is designed to be adaptable to other university environments.*

*Mots-clés : modèle de maturité, gouvernance des données, enseignement supérieur*

*KEYWORDS: maturity model, data governance, higher education*

## 1. Introduction

Dans le contexte d'une production très rapide des données[1], la question de leur gouvernance traverse l'ensemble des organisations. Si cette problématique a été principalement étudiée dans un contexte entrepreneurial (Jimenez et al., 2019) où une gestion optimale des données sert en priorité des objectifs économiques, elle a

[1] https://www.statista.com/statistics/871513/worldwide-data-created/

moins fait l'objet d'études dans un cadre universitaire. Or, l'absence de gouvernance expose indifféremment les organisations à de nombreux risques d'ordre réglementaire, sécuritaire et logistique qui entravent la bonne gestion des données et amènent à questionner leur fiabilité et légitimité (Verdier, 2015 ; Al-Ruithe, Benkhelifa et Hameed, 2016 ; Kremser et Brunauer, 2019) ; ce que Borgman et Brand (2022) ont nommé une « taxe invisible » sur l'efficacité. Le monde universitaire anglophone s'est emparé de cette thématique en inscrivant la gouvernance des données dans leur fonctionnement depuis une dizaine d'années (Jim et Chang, 2018). En France, les initiatives visant la mise en place d'une gouvernance des données sont assez confidentielles[2]. De plus, les modes d'expression de la gouvernance universitaire en France, bien plus décentralisée que celles du monde anglophone (Melançon, Pinède et Verdi, 2024), ajoutent de la difficulté à une situation naturellement compliquée. En effet, la mise en place d'une gouvernance des données est complexe et soulève des problématiques diverses. En amont de la nécessité de les résoudre se pose celle de les identifier. Cet article s'inscrit dans le cadre d'une recherche action, celle du projet GouD[3], afin de proposer un modèle de maturité pour permettre aux universités de disposer d'un outil concourant d'une part à identifier et analyser les mécanismes de la gouvernance des données et, d'autre part, à évaluer le niveau d'implication des acteurs et de l'organisation. Bien qu'adapté à un contexte local particulier, à savoir les projets ACT de l'université de Bordeaux, il a pour vocation d'être adaptable à d'autres contextes universitaires.

## 2. La gouvernance des données en contexte universitaire

Dans nos précédents travaux (Melançon et Pinède, 2023 ; Verdi, 2023a), nous avions défini la gouvernance des données comme l'exercice de l'autorité et du contrôle sur la gestion des données par l'institution d'un système de normes et de procédures (Plotkin, 2013). Celle-ci vise une exploitation optimale des données, notamment grâce à une politique générale d'ouverture des données (Gegenhuber et al., 2023), afin d'aiguiller les décisions de l'organisation (Janssen, 2020). Dans ce cadre, la gouvernance doit s'assurer du respect de la conformité de ses principes (Loshin, 2008), ce qui est souvent englobé dans le vocable anglais de *compliance* (Fu et al., 2011). Ce système de principes et de règles, intégré dans une stratégie globale de long terme (Weber, Otto et Österle, 2009 ; Waller, 2020), induit la constitution d'une culture commune sur les données où ces dernières sont valorisées et perçues comme porteuses de valeur pour les acteurs de l'université. Son émergence est grandement liée à la culture des données (*data literacy)* diffusée au sein de l'université. Cette dernière comprend à la fois des connaissances théoriques et techniques mais intègre également un aspect de formation (Verdi, 2023b). Elle est

[2] Le seul véritable exemple est celui de l'Université Côte d'Azur (voir AMUE, 2022).

[3] Le projet GouD est l'un des projets du programme ACT (*Augmented university for Campus and world Transition*) de l'université de Bordeaux qui mène un travail prospectif sur la gouvernance des données et examine l'état de l'art sur les questions de gouvernance pour cerner les particularités du contexte universitaire français et *in fine* proposer un modèle de gouvernance spécifiques aux établissements ESRI.

essentielle à la gouvernance en permettant « l'enculturation » des acteurs qui, au-delà d'une simple acculturation, favorise le mécanisme d'accomplissement du changement et amène à une acceptation plus naturelle de ses principes (Herskovits, 1952). Culture des données et gouvernance des données sont donc indissociables et forment deux « blocs constitutifs importants du socle de connaissances des professionnels de l'information impliqués dans le soutien des recherches intensives des données »[4] (Koltay, 2016).

La gouvernance englobe ou est englobée dans un certain nombre de modalités. En premier lieu, la gestion des données (*data management*) : au contraire de la gouvernance qui institue les principes, les règles et les valeurs ainsi que la répartition des rôles des acteurs qui la composent, la gestion des données a pour but la mise en place effective de ceux-ci (Khatri et Brown, 2010 ; Alhassan, Sammon et Daly, 2018 ; Rafal et Girard, 2023). Cette gestion des données se confond souvent dans la littérature avec l'intendance des données (*data stewardship*) qui se présente également comme la facette opérationnelle de la gouvernance (Plotkin, 2013). Dans les deux cas, qu'il s'agisse de la gestion des données ou de l'intendance des données, l'objectif poursuivi est le même : s'assurer de la qualité des données, comprise dans l'activité de la gestion de la qualité des données (*data quality management*) (Wende et Otto, 2007). Celle-ci comprend la conformité de leur complétude, de leur consistance, de leur précision, de leur pertinence, de leur interprétabilité, de leur réutilisabilité ou encore de leur rapidité d'obtention (Pinino, Lee et Yang, 2002 ; Cheong et Chang, 2007 ; Brous, Janssen et Herder, 2016). Cette gestion de la qualité, au-delà de s'assurer de la bonne facture des données, doit prend en compte la sécurité de ces dernières (*data security*) pour éviter toute fuite ou perte (pouvant porter atteinte à la vie privée des acteurs par exemple), ce qui suppose une accréditation des personnels (Benfeldt, Persson et Madsen, 2020) et la mise en place d'un ensemble de procédures de protection à court et long terme. La difficulté réside ici dans la recherche d'un bon équilibre entre l'accès et le contrôle des données pour ne pas enrayer le fonctionnement de la gouvernance (Rafal et Girard, 2023).

Pour assurer ce fonctionnement, plusieurs typologies d'acteurs sont intégrées à la gouvernance. Nous avons pu lister (1) les administrateurs (*data trustees*) qui sont les responsables de la conformité du respect des règles de la gouvernance pour garantir l'intégrité et l'utilité des données ; (2) les propriétaires (*data owners*) à l'origine de la conception des jeux de données et peuvent être tout aussi bien des personnels internes (ex : enseignant-chercheur) qu'externes à l'université (ex : une institution comme l'*INSEE*) ; (3) les utilisateurs (*data users*) qui s'emparent des jeux de données pour des usages spécifiques et qui n'en sont pas nécessairement les destinataires originels ; 4) les intendants (*data stewards*) ou managers (*data managers*) qui sont des experts disciplinaires donnant des conseils sur les différents aspects liés à la gestion des données (Teperek et al., 2018). Il est à noter que les personnes occupant les rôles (1) et (2) ne peuvent pas toutes être incluses dans la

[4] « *Data governance and data literacy are two important building blocks in the knowledge base of information professionals involved in supporting data-intensive research, and both address data quality and research data management* ».

gouvernance pour des raisons pratiques : désigner des représentants pourrait être une solution plus simple pour recueillir leurs contributions.

Dans le monde universitaire français, ces rôles ne sont pas nécessairement constitués comme tels et un acteur peut assurer à lui seul plusieurs de ces fonctions comme dans le cas d'un délégué à la protection des données (DPO). Néanmoins, cette conceptualisation est utile pour délimiter un premier périmètre d'étude. Avec l'aval des instances dirigeantes, ces acteurs peuvent ensuite se constituer en comité, pourvu ou non d'un pouvoir décisionnel. Dans les universités nord-américaines, un même modèle de gouvernance, adapté du DMBOK[5], est visible : celui-ci est vertical, piloté en premier lieu par le *provost* (que nous pourrions rapprocher du rôle de Directeur Général des Services) qui prend part, voire constitue, un comité de la gouvernance des données (*data governance board* ou *council*) intégrant des acteurs issus de l'ensemble des services de l'université[6] qui élaborent et supervisent l'application des mécanismes. Au contraire de cette approche descendante qui transforme la gouvernance en un organe de contrôle, nous avions souligné dans Melançon et Pinède (2023) que l'approche « non-intrusive » de Seiner (2014) mettant la gouvernance au service des projets semblait plus adaptée au contexte local français. Dans le cadre de l'université de Bordeaux, cette implémentation débute précisément par un ensemble d'initiatives limitées à l'écosystème des projets ACT que nous allons maintenant présenter.

### 3. Le contexte bordelais : le projet ACT

Débuté en 2021, le projet *Augmented university for Campus and world Transition* (ACT) est un projet ANR ayant pour mission « de développer, tester, valider et diffuser de nouvelles façons d'aborder les grands problèmes environnementaux, sociaux et de transition économique grâce à la transformation des campus de l'université en un vaste laboratoire vivant »[7] . Il comprend plus de 24 projets, intégrés dans des *living labs,* liés aux questions de transition (écologique, économique et numérique) où interagissent un grand nombre d'acteurs et dans lesquels les productions de données s'insèrent dans des visées différentes. Nous pouvons citer à titre d'exemple les projets « PRISME » manipulant les données de santé d'étudiants, « Datacampus » travaillant sur les données de mobilité du campus, ou encore « Forêt Urbaine » analysant les données issues des réponses de la forêt urbaine au changement climatique. En surplomb se greffe le projet « Datalab », débuté en 2023, dont l'objectif est de développer, à l'échelle des projets ACT, des actions de stratégie data, de gouvernance des données et de développement des

[5] Le DMBOK est une publication de la DAMA International qui décrit les mécanismes de la gouvernance des données dans un contexte entrepreneurial.

[6] Nous retrouvons ce modèle dans de nombreuses universités nord-américaines comme celles d'East Carolina, de Madison, du Maine, de San Francisco ou de Villanova, mais également sous une forme similaire dans les universités de Toronto (Canada) et de Londres (Royaume-Uni).

[7] https://www.u-bordeaux.fr/universite/notre-strategie/projets-institutionnels/act-pour-un-campus-experimental

usages de la donnée (Blanchard, 2023). La volonté d'instauration d'une gouvernance des données est néanmoins plus ancienne : elle s'est faite jour officiellement avec la création d'un COPIL « Gouvernance des données et des documents d'activité publics de l'établissement » (GDDP) en 2017 sous l'impulsion de la Direction des archives universitaires, au sein de laquelle une réflexion et des travaux ont menés à la production d'une note interne de contexte sur la gouvernance des données en 2021 et d'un article d'Anne Pletinckx, directrice des Archives universitaires (Pletinckx, 2022), qui proposent tous deux une description des caractéristiques de la gouvernance des données et qui soulignent les prémisses de l'intérêt du sujet au sein de l'université. La gouvernance n'en est actuellement qu'à un stade embryonnaire, entamant depuis peu la poursuite d'initiatives telles que la réalisation de Plans de Gestion des données (PGD), la constitution en décembre 2023 d'un groupe accueillant un certain nombre d'acteurs internes et externes à l'université de Bordeaux contribuant à la définition des principes de la gouvernance et dernièrement avec la création en février 2024 d'un service 3D (Données, décisionnel, *datalab*) ayant en charge l'animation de la gouvernance des données et la diffusion d'une culture des données au sein de l'université. En tant que projet de recherche-action s'appuyant sur une démarche empirique et inductive, GouD utilise l'ensemble des projets ACT précités comme terrain d'expérimentation. Le modèle de maturité, issu de notre lecture de la littérature mais également des observations issues des projets ACT, s'inscrit ici dans une analyse de long terme de l'élaboration de la gouvernance.

## 4. Le modèle de maturité

Les objectifs poursuivis par notre modèle de maturité sont les suivants : clarifier les caractéristiques générales de la gouvernance des données et pouvoir les analyser à travers le prisme de trois facettes. Pour chaque facette, des composantes seront décrites et des sources permettant de les analyser seront présentées. Le tout pour aider à la mise en place d'une gouvernance des données ou pour en améliorer une déjà existante. Notre modèle de maturité doit donc permettre de poser un constat sur une situation en cours, à différents stades, et anticiper son évolution.

Selon la littérature, un modèle de maturité sert d'outil de diagnostic (Kohlegger, Maier & Thalmann, 2009) ayant été testé dans un contexte spécifique (Wilkinson, 2014) et comprenant deux perspectives potentielles : un cycle de vie ou un potentiel de performance (McBride, 2010). Il peut être construit selon un modèle descriptif, prescriptif ou comparatif (Poeppelbuss et Roeglinger, 2011) et intègre des niveaux de maturité pour un ensemble d'items comme des processus, des rôles ou une organisation par exemple. Ces niveaux représentent une ligne d'évolution anticipée ou désirée (Becker, 2009). Notre modèle doit permettre de déterminer un niveau de maturité de la gouvernance des données mise en place au sein des projets ACT de l'université de Bordeaux. A ce titre, il doit intégrer dans son analyse l'ensemble des acteurs ayant une interaction avec les données employées au sein de l'université : les acteurs des projets ACT mais également ceux des directions de services (ex : la direction des archives universitaires), voire les partenaires extérieurs (ex : la

Métropole de Bordeaux). Les concepts précités (ex : *data management, data quality, data literacy,* etc.) sont ici ré-agencés pour proposer une grille de lecture adaptée à un contexte français.

La maturité de la gouvernance des données n'est que peu définie dans la littérature, à l'exception de Marchildon et al. (2018) et de Gupta et Cannon (2020). Selon ces auteurs, celle-ci renvoie au niveau de conception et de déploiement de règles et de principes visant à optimiser la gestion des données sur leur cycle de vie, à savoir leur détection, leur production, leur collecte, leur usage, leur partage, leur stockage, leur archivage et leur destruction. Pour être mesurable, elle nécessite l'analyse de deux critères : l'existence ou non d'une variable (ex : les moyens humains) et la capacité de l'université à pouvoir identifier ces variables (ex : les données employées dans un laboratoire). Comme tout modèle de maturité, le nôtre intègre des niveaux qui explicitent la formalisation, la généralisation et l'adhésion aux principes de la gouvernance. Dans la littérature, ils sont généralement au nombre de cinq et leurs intitulés diffèrent selon les sources (Soares, 2013 ; Seiner, 2014 ; Merkus, 2015 ; Lee et al., 2019). Carretero et al. (2016) citent un niveau 0 pour signifier une absence totale d'initiatives. Nous avons fait le choix de six niveaux de maturité : un niveau 0 (inexistante), un niveau 1 (en émergence), un niveau 2 (en développement), un niveau 3 (déploiement local), un niveau 4 (déploiement global) et un niveau 5 (mature). Ces niveaux se déclinent selon trois facettes : les données, les acteurs et l'organisation. Leur sélection s'est faite à partir de la littérature (scientifique et professionnelle) décrivant les caractéristiques de la gouvernance des données mais également à partir d'autres facettes mises en avant par les auteurs (comme les personnes, les normes ou encore les technologies) que nous avons remaniées ainsi que des difficultés rencontrées dans l'élaboration de la gouvernance. Ces deux derniers points, bien que n'étant pas originellement intégrés dans une logique d'analyse de maturité, ont toutefois permis d'affiner notre proposition.

Plusieurs auteurs et rapports ont contribué à lister les freins potentiels et les défis à relever pour pouvoir construire et asseoir une gouvernance des données (DAMA International, 2009 ; Benfeldt, 2017 ; Mahanti, 2021 ; Okoro, 2021). Y sont cités les problèmes (1) de compréhension commune sur ce que recouvre la gouvernance des données, (2) de logistique vis-à-vis de la mise en place et de la conformité de son application, (3) de conflits entre acteurs et/ou services dus à des questions de partage de pouvoir, de périmètres d'action ou encore de visions concurrentes sur les données et leur traitement, (4) de refus du changement découlant d'un rejet des règles et principes de la gouvernance, (5) d'absence de compétences et de moyens (matériels, financiers ou humains) pour traiter les nécessités de la gouvernance. Au point (5), l'absence est due soit à une non-spécialisation des acteurs, soit à une incapacité de les définir et *in extenso* de recruter les personnes compétentes, soit à un désinvestissement de cette problématique par les personnes ayant un pouvoir décisionnel, notamment celles en charge du budget de l'établissement. Ces points sont transversaux et doivent être déduits de l'analyse de chaque composante de notre modèle de maturité, en particulier la non-adhésion aux principes de la gouvernance.

La première facette traite des données. On lui dénombre trois sous-facettes : (1) le capital des données, c'est-à-dire l'ensemble des données existantes et les moyens disponibles pour les recenser et les gérer, (2) les règles et principes encadrant cette gestion et (3) les formations proposées pour établir une culture commune des données.

*Tableau 2: la facette des données*

| Sous-facettes | Composantes | Description des composantes |
|---|---|---|
| Le capital | Les données recensées | L'ensemble des données utilisées dans différents projets ou contextes. Les caractéristiques d'un ensemble de données doivent être documentées de manière exhaustive en précisant leur formats, leurs métadonnées, etc. |
| | Les outils de mesure | Les livrables permettant de lister et surveiller les flux de données (ex : un plan de gestion de données, un *dashboard* ou un catalogue de données). |
| | Les moyens | Les caractéristiques des moyens humains (les tâches et les rôles), structurels (ex : les serveurs, les logiciels, les locaux dédiés) et financiers. |
| Les règles et les principes | Le cycle de vie des données | Les règles et principes encadrant l'ensemble du cycle de vie des données, à savoir l'identification, la collecte, la production, le stockage, l'archivage et la destruction des données (ex : intégrés dans un arbre décisionnel). Si rendus obligatoires, les logiciels et technologies choisis doivent ici être précisés et décris. Du fait du grand nombre de tâches que cette composante implique, elles peuvent être traitées séparément (dans le cadre d'un audit par exemple). |
| | La qualité | L'ensemble des règles et principes employés pour s'assurer de la pertinence, de la complétude, de la conformité, de l'intégrité, de la fraîcheur et de la cohérence des données (ex : principes *FAIR*, norme des métadonnées, etc.). |
| | La sécurité | L'ensemble des protections matérielles et cyber ainsi que la gestion des accès aux données. |
| | La documentation | Les livrables listant l'ensemble des règles et des principes de gestion des données. |
| Les formations | Les formats | Le déroulement des formations (ex : la temporalité, les activités proposées, etc.) et les sujets qu'elles abordent (ex : la science ouverte). |
| | Les publics visés | Les types de publics (ex: les étudiants, les directions universitaires, etc.) et les raisons de ce ciblage (ex : amélioration de compétences, développement de nouvelles connaissances, etc.). |
| | Les formateurs | Le profil des formateurs (ex : est-ce un acteur interne / externe, quelle est la nature de son poste, etc.) et la raison de leur recrutement (ex : compétences uniquement détenues par ces derniers, partenariat historique avec l'université, etc.). |

La seconde facette traite des acteurs. Son objectif est de souligner le rôle humain en prenant en considération l'autonomie des acteurs et leur aptitude à accepter ou non le changement. Devront ici être soulevés leurs rôles respectifs, leurs méthodes de travail, leurs modes de collaboration ainsi que le degré de leur culture des données. Il s'agit donc ici d'une enquête portant à la fois sur les pratiques effectives et sur les imaginaires des acteurs.

*Tableau 3: la facette des acteurs*

| Sous-facettes | Composantes | Description des composantes |
|---|---|---|
| Les typologies | Les administrateurs | Les acteurs en charge du respect de l'application des principes et règles de traitement des données. |
| | Les producteurs | Les acteurs en charge de la production des données. |
| | Les utilisateurs | Les acteurs employant les données mises à disposition par l'université. |
| | Les intendants ou managers | Les acteurs conseillant et supervisant la gestion des données. |
| Les usages | Les contextes et les objectifs | La description des contextes de production et de traitement des données ainsi que les objectifs que servent les données. |
| | Le traitement | La description du traitement des données sur l'ensemble de leur cycle de vie (ex : quel acteur traite quelle donnée, selon quelle temporalité, quels matériels, dans quel objectif, etc.). |
| | La conformité | La capacité de la gouvernance d'intégrer les différents cadres réglementaires (ex : *RGPD*), de pouvoir dialoguer avec d'autres instances ayant également développé des règles (ex : comité éthique, comité de santé, etc.) et de s'assurer du respect des règles établies. |
| Les collaborations | Les modalités de collaboration | La manière dont les acteurs collaborent entre eux (ex : leurs méthodologies, leurs connaissances et représentations respectives). |
| La culture des données | Les niveaux de formation | Les connaissances et compétences des acteurs sur les données et leur écosystème dont fait partie la gouvernance des données. |
| | Les représentations | Les visions et imaginaires des acteurs sur les données et leur écosystème dont fait partie la gouvernance. |

La troisième facette traite de l'organisation et se réfère à l'élaboration et la supervision de la gouvernance. Elle liste l'ensemble des éléments à prendre en considération pour garantir sa pérennité. Elle contient (1) la stratégie qui vise à établir et à faire connaître la vision de la gouvernance défendue par l'université

auprès des acteurs qui la composent, (2) la forme qui la régit et (3) les collaborations entretenues entre acteurs internes et externes à l'université.

*Tableau 4: la facette de l'organisation*

| Sous-facettes | Composantes | Description des composantes |
|---|---|---|
| La stratégie | La vision stratégique | Les valeurs et les objectifs globaux de la gouvernance ainsi que la manière dont elle est reçue et interprétée par les acteurs de l'université. |
| | La communication | Les caractéristiques des stratégies de communication (ex : les acteurs qui en ont la charge, la temporalité des actions de communication, les sujets abordés, etc.) et leur réception par les acteurs (ex : adhésion, rejet, etc.). |
| La forme | La coordination de la gouvernance | Précise la manière dont la gouvernance est coordonnée (centralisation, décentralisation ou subsidiarité) et les raisons ayant conduit à ce choix. |
| | L'attribution des rôles | Désigne les mécanismes d'attribution des rôles aux acteurs pour traiter un point spécifique de la gouvernance. |
| Les collaborations | Les collaborations internes | Les acteurs internes à l'université (ex : directions de services) et les modalités des collaborations (ex : rôle consultatif permanent, temporaire, etc.) |
| | Les collaborations externes | Les acteurs externes à l'université et les modalités des collaborations. |

Au même titre que les mécanismes de la gouvernance décrits par Abraham et al. (2019), notre modèle se veut interactionniste : chaque facette a une influence sur les autres et aucune n'est prévalente. Toutes doivent être développées pour que la gouvernance des données atteigne un niveau de maturité acceptable, déterminé selon les besoins et objectifs de l'université.

L'ensemble des points précédents doit permettre ainsi de déterminer un niveau de maturité pour chacune des facettes à un instant t. En effet, de nombreux changements internes (budgétaires, matériels, logistiques et humains) ont lieu dans le temps, en particulier le *turnover* fréquent des personnels, auxquels s'ajoutent d'autres changements externes comme l'évolution des réglementations. Le niveau de maturité mesuré n'est donc pas définitif et devra être à nouveau analysé.

Dans notre tableau récapitulatif des niveaux par facettes, les cases ne doivent pas se lire de manière indépendante, les composantes d'une facette influant directement celles d'une autre facette. A titre d'exemple, sans actions de communication, les représentations des acteurs sur la gouvernance n'évolueront pas.

Ainsi, dans cette conception, les niveaux ne sont pas homogènes : une facette pourra par exemple être à un niveau 2 tandis qu'une autre sera à un niveau 0. L'évaluation globale du niveau de maturité est appréciative, dépendante des contextes spécifiques des organisations.

*Tableau 5: les niveaux de maturité par facette*

| Niveaux / facettes | Les données | Les acteurs | L'organisation |
|---|---|---|---|
| Niveau 0 Inexistante | Les données ne sont pas identifiées, ni traitées selon des principes et des règles unifiés. Aucune formation aux données n'est proposée à ce niveau. | Les acteurs ne sont pas identifiés. Leur culture des données est trop peu développée pour qu'ils puissent avoir une vision claire de ce que recouvre la gouvernance. | La gouvernance n'est ni définie, ni déployée. |
| Niveau 1 En émergence | Une volonté d'identifier et de gérer les données selon des principes et des règles unifiés émerge. Des formations aux données ne sont pas encore proposées aux acteurs. | De rares acteurs ainsi que les modalités de leur collaboration et de leur usage des données sont identifiés. Ils ne comprennent pas complètement ce que recouvre la gouvernance. | Le besoin d'une gouvernance émerge. Aucune stratégie de communication n'est encore élaborée. |
| Niveau 2 En développement | Quelques données commencent à être identifiées. L'acculturation aux données passe par des formations externes selon les opportunités rencontrées par les acteurs. | Quelques acteurs ainsi que les modalités de leur collaboration et de leur usage des données sont identifiés. Ils commencent à comprendre ce que recouvre la gouvernance. | La gouvernance est cours d'élaboration. Quelques actions de communication sont potentiellement entreprises pour en discuter. |
| Niveau 3 Déploiement local | Une partie seulement des données est identifiée et est gérée selon les principes et les règles de la gouvernance. Des formations aux données sont proposées ponctuellement par la gouvernance pour soutenir les acteurs. | Une minorité d'acteurs ainsi que les modalités de leur collaboration et de leur usage des données sont identifiés. Leur culture des données développée leur permet de comprendre ce que recouvre la gouvernance. | La gouvernance est déployée à un niveau local, soutenue par une première stratégie de communication qui met à disposition une documentation détaillant ses principes et ses visées. |

| | | | |
|---|---|---|---|
| Niveau 4 Déploiement global | La majorité des données sont identifiées. Celles-ci sont gérées selon les principes et des règles de la gouvernance. Des formations sont mises en place de manière régulière par la gouvernance pour soutenir les acteurs. | La majorité des acteurs ainsi que les modalités de leur collaboration et de leur usage des données sont identifiés. Leur culture des données développée leur permet de comprendre ce que recouvre la gouvernance. | La gouvernance est déployée à un niveau global et une stratégie de communication est désormais entreprise de manière constante. Elle rencontre néanmoins encore des difficultés d'application. |
| Niveau 5 Mature | L'ensemble des données sont identifiées et gérées selon les principes et règles de la gouvernance. Des formations sur les données sont proposées par la gouvernance de manière constante pour répondre aux besoins des acteurs. | La totalité des acteurs ainsi que les modalités de leur collaboration et de leur usage des données sont identifiés. Ils comprennent et adhèrent aux principes et règles de la gouvernance. | La gouvernance est déployée à l'ensemble de l'université et sa comitologie est étendue aux partenaires de son organisation. Elle est robuste et capable de s'adapter à de nouveaux contextes. |

Comme précisé en introduction, l'une des plus grandes difficultés réside dans l'identification d'éléments constitutifs d'une gouvernance, en particulier lorsque ce terme n'est pas employé : soit parce qu'il ne recouvre pas pour les acteurs une volonté de stratégie globale, soit parce qu'il est tout simplement délaissé pour un autre plus représentatif des activités entreprises comme la science ouverte. Or, des initiatives autour des données peuvent constituer le terreau de la gouvernance et doivent en ce sens faire l'objet d'analyses. C'est le cas des ateliers de la donnée qui se positionnent « comme le point d'entrée en proximité des équipes de recherche sur toute nature de besoin relatif à la donnée »[8] et dont la constitution en réseau facilite les collaborations et l'instauration d'une vision commune sur les données. En nous basant sur la littérature et nos retours de terrain, nous proposons ici trois grandes sources d'analyses pour abonder notre modèle de maturité pour la gouvernance des données.

D'une part, les sources documentaires qui listent les principes et les règles sur le traitement des données. Celles-ci qui s'incarnent dans de nombreux formats. L'on peut ici citer les chartes, référentiels et règlements intérieurs, les arbres décisionnels, les comptes rendus de réunions, les contenus web ainsi que toute documentation

[8] https://www.ouvrirlascience.fr/wp-content/uploads/2021/10/2021.10.11_AMI_Ateliers-de-la-donne%CC%81e.pdf

annexe qui permettent de déceler les règles en application au sein de l'université et les visions sous-jacentes dans la gestion des données. Assister à leur élaboration apporte une possibilité supplémentaire d'analyser l'évolution des pensées et les modes d'interaction entre acteurs.

Nous recommandons en particulier l'étude des Plans de Gestion des Données (PGD). De par leur description des jeux de données, ils offrent un panorama complet sur le capital de l'université. De plus, ils permettent également d'identifier les acteurs des projets et leurs rôles respectifs. Bien que centrés autour des données de la recherche, ils peuvent toutefois être déclinés pour d'autres types de données et couvrir un plus large ensemble de ressources. Leur élaboration rend possible le questionnement des acteurs pour obtenir des détails plus précis sur leurs méthodes de travail et sur leurs interactions. Bien qu'étant une source documentaire, la singularité de leurs caractéristiques en font à notre sens un élément d'analyse dissocié des autres sources documentaires.

D'autre part, les sources d'enquêtes des acteurs qui peuvent être réalisées soit avec des questionnaires, soit avec des entretiens. A titre d'exemple, Marchildon et al. (2018) ont proposé un questionnaire à choix multiples constitué de 11 items composés de 72 questions permettant d'analyser chaque aspect d'une gouvernance existante. Si les questionnaires permettent des analyses quantitatives immédiates, les entretiens offrent eux un approfondissement des points peu développés dans les documentations précitées. C'est particulièrement le cas de la culture des acteurs qui ne peut émerger que par ce type d'enquête. Cette culture, par-delà les compétences, connaissances et représentations, induit des pratiques spécifiques, notamment dans le cadre d'une discipline ou d'un service instauré depuis longtemps.

Déceler ces pratiques et contraintes est ici primordial pour adapter la gouvernance aux pratiques locales et éviter qu'elle ne soit perçue comme un doublon incohérent ou comme une charge de travail supplémentaire par les acteurs. Savoir qui interroger relève d'une autre difficulté, en particulier sans historique des précédentes actions réalisées. Dans notre cas, les acteurs liés aux projets ACT d'une part et les acteurs liés à la constitution de la gouvernance de l'université de Bordeaux d'autre part ont pu être identifiés en amont de ce travail. Mais nous gardons à l'esprit que d'autres acteurs ayant un intérêt à la gouvernance et ne s'étant pas encore manifestés peuvent avoir mis en place un certain nombre d'initiatives en ce sens, nécessitant alors le déploiement d'une nouvelle vague d'entretiens. Ces trois sources peuvent être sollicitées pour l'ensemble des facettes mais certaines donneront des réponses plus précises que d'autres, c'est pourquoi nous recommandons ci-dessous les plus appropriées.

*Tableau 6: les sources recommandées par facettes*

| Facettes | Sous-facettes | Sources documentaires | Plan de gestion des données | Sources d'enquêtes |
|---|---|---|---|---|
| Les données | Le capital | limitée | ✓ | optionnelle |
| | Les règles et les principes | ✓ | ✓ | optionnelle |
| | Les formations | ✓ | hors périmètre | ✓ |
| Les acteurs | Les typologies | ✓ | ✓ | ✓ |
| | Les usages | ✓ | ✓ | optionnelle |
| | Les modalités de collaboration | ✓ | ✓ | ✓ |
| | La culture des données | limitée | peu adaptée | ✓ |
| L'organisation | La stratégie | ✓ | hors périmètre | ✓ |
| | La forme | ✓ | hors périmètre | ✓ |
| | Les collaborations | ✓ | hors périmètre | ✓ |

## 5. Conclusion

L'édification et la pérennisation d'une gouvernance des données est un processus long et complexe, en particulier en contexte universitaire. A l'heure où la gestion des données revêt une importance cruciale, il devient nécessaire d'en maîtriser les tenants et aboutissants. La synthèse de la littérature a eu pour objet de proposer une caractérisation de la gouvernance et appelle ici une validation. Le modèle de maturité présenté est actuellement testé dans le contexte d'ACT. La lecture de la documentation, dont les PGD, et la réalisation d'enquêtes entreprises tout au long de l'année 2024 devront permettre de décrire les caractéristiques de la gouvernance des données de l'université de Bordeaux et d'en mesurer le niveau de maturité. Elles pourront amener à une évolution des facettes et de leurs composantes afin de mieux cerner les particularités de la gouvernance des données en contexte universitaire. Ceci fera l'objet de futures publications.

*Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme d'Investissements d'avenir portant la référence ANR-20-IDES-0001*

## Bibliographie

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance : A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities : A comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management*, *31*(2), 300316. https://doi.org/10.1108/JEIM-01-2017-0007

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2016). A Conceptual Framework for Designing Data Governance for Cloud Computing. *Procedia Computer Science*, *94*, 160167. https://doi.org/10.1016/j.procs.2016.08.025

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management. *Business & Information Systems Engineering*, *1*(3), 213222. https://doi.org/10.1007/s12599-009-0044-5

Benfeldt, O. (2017). A Comprehensive Review of Data Governance Literature. *Selected Papers of the IRIS, Issue Nr 8 (2017)*, *3*, 120133. https://aisel.aisnet.org/iris2017/3

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299313. https://doi.org/10.1007/s10796-019-09923-z

Blanchard, A. (2023). Enjeux de la donnée universitaire, de sa collecte à son exploitation. *Les data de l'ESR : l'enjeu de la maîtrise des données. Risques et outils pour l'audit et le pilotage.* https://hal.science/hal-04289500

Borgman, C. L., & Brand, A. (2022). Data blind : Universities lag in capturing and exploiting data. *Science*, *378*(6626), 12781281. https://doi.org/10.1126/science.add2734

Brous, P., Janssen, M., & Herder, P. (2016). Coordinating Data-Driven Decision-Making in Public Asset Management Organizations : A Quasi-Experiment for Assessing the Impact of Data Governance on Asset Management Decision Making. *Social Media : The Good, the Bad, and the Ugly*. Springer International Publishing, p. 573583

Carretero, A., Freitas, A., Cruz-Correia, R., & Piattini, M. (2016, janvier 1). *A case study on assessing the organizational maturity of data management, data quality management and data governance by means of MAMD*.

Cheong, L. K., & Chang, V. (2007). *The Need for Data Governance : A Case Study*.

DAMA International. (2009). *The DAMA Guide to the Data Management Body of Knowledge —DAMA-DMBOK*. Technics Publications, LLC.

Fu, X., Wojak, A., Neagu, D., Ridley, M., & Travis, K. (2011). Data governance in predictive toxicology : A review. *Journal of Cheminformatics*, *3*(1), 24. https://doi.org/10.1186/1758-2946-3-24

Gegenhuber, T., Mair, J., Lührsen, R., & Thäter, L. (2023). Orchestrating distributed data governance in open social innovation. *Information and Organization*, *33*(1), 100453. https://doi.org/10.1016/j.infoandorg.2023.100453

Gupta, U., Cannon, S. (2020). Data Governance Maturity Models. In *A Practitioner's Guide to Data Governance* (p. 143165). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78973-567-320201007

Herskovits, M. J. (1952). *Les Bases de l'anthropologie culturelle*. https://unesdoc.unesco.org/ark:/48223/pf0000244386

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance : Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Jim, C. K., & Chang, H. (2018). The current state of data governance in higher education. *Proceedings of the Association for Information Science and Technology*, *55*(1), 198-206. https://doi.org/10.1002/pra2.2018.14505501022

Jimenez, L. M., Polo, J. A., & Duarte, N. A. (2019). Overview of Data Governance in Business Contexts. *IOP Conference Series: Materials Science and Engineering*, *519*(1), 012023. https://doi.org/10.1088/1757-899X/519/1/012023

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148152. https://doi.org/10.1145/1629175.1629210

Kohlegger, M., Maier, R., & Thalmann, S. (2009). *Understanding Maturity Models. Results of a Structured Content Analysis*. I-KNOW '09 and I-SEMANTICS '09. https://www.researchgate.net/publication/215312013_Understanding_Maturity_Models_Results_of_a_Structured_Content_Analysis

Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal*, *42*(4), 303312. https://doi.org/10.1177/0340035216672238

Kremser, W., & Brunauer, R. (2019). Do we have a Data Culture? In P. Haber, T. Lampoltshammer, & M. Mayr (Éds.), *Data Science – Analytics and Applications* (p. 8387). Springer Fachmedien. https://doi.org/10.1007/978-3-658-27495-5_11

Lee, Y., Park, W., Shin, D., & Won, Y. (2019). A Study on Data Governance Maturity Model and Total Process for the Personal Data Use and Protection. *Journal of the Korea Institute of Information Security & Cryptology*, *29*(5), 11171132. https://doi.org/10.13089/JKIISC.2019.29.5.1117

Loshin, D. (2008). Master Data Management. Morgan Kaufmann.

Mahanti, R. (2021). *Data Governance Success : Growing and Sustaining Data Governance*. Springer. https://doi.org/10.1007/978-981-16-5086-4

Marchildon, P., Bourdeau, S., Hadaya, P., & Labissière, A. (2018). Data governance maturity assessment tool : A design science approach. *Projectics / Proyéctica / Projectique*, *20*(2), 155193. https://doi.org/10.3917/proj.020.0155

McBride, T. (2010). Organisational theory perspective on process capability measurement scales. *Journal of Software Maintenance and Evolution: Research and Practice*, *22*(4), 243254. https://doi.org/10.1002/spip.440

Melançon, G., Pinède, N., et Verdi, U. (2024). Redefining Data Governance: Insights from the French University System. *Actes du colloque ICEIS 2024*, Angers.

Melançon, G., & Pinède, N. (2023). Gouvernance des données et intelligibilité : Une approche méthodologique en contexte universitaire. *Communication & Organisation*, *64*(2), 6781. https://doi.org/10.4000/communicationorganisation/12664

Merkus, J. (2015). *Data Governance Maturity Model*. https://doi.org/10.13140/RG.2.2.19274.16321

Okoro, R. (2021). Proposed Data Governance Framework for Small and Medium Scale Enterprises (SMEs). *All Graduate Theses, Dissertations, and Other Capstone Projects*. https://cornerstone.lib.mnsu.edu/etds/1126

Pinino, L. L., Lee, W. W., & Yang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4).

Pletinckx, A. (2022). De la gouvernance des données publiques. *Archivistes ! La lettre de l'Association des archivistes français*, *140*, 30-31.

Plotkin, D. (2013). *Data Stewardship : An Actionable Guide to Effective Data Management and Data Governance*. Newnes.

Poeppelbuss, J., & Roeglinger, M. (2011). *What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management*. 19th European Conference on Information Systems, ECIS 2011.

Rafal, O., & Girard, D. (2023). *Maîtriser la #data : Un enjeu majeur de 2023*. WENVISION. https://www.wenvision.com/maitriser-la-data-un-enjeu-majeur-de-2023/

Seiner, R. S. (2014). *Non-Invasive Data Governance : The Path of Least Resistance and Greatest Success*. Technics Publications.

Soares, F. S. F., & De Lemos Meira, S. R. (2013). *An Agile Maturity Model for Software Development Organizations*. 4.

Teperek, M., Cruz, M. J., Verbakel, E., Böhmer, J., & Dunning, A. (2018). Data Stewardship Addressing Disciplinary Data Management Needs. *International Journal of Digital Curation*, *13*(1), Article 1. https://doi.org/10.2218/ijdc.v13i1.604

Verdi, U. (2023a). L'éthique des données dans les chartes éthiques des collectivités territoriales. *Communication et organisation, 64*.

Verdi, U. (2023b). Quelle(s) réponse(s) à l'enjeu d'acculturation aux données ? Un état de l'art des caractéristiques de la data literacy. *Revue française des sciences de l'information et de la communication*, *26.*

Verdier, H. (2015). *Administrateur général des données—Rapport au Premier ministre sur la gouvernance de la donnée 2015 : Les données au service de la transformation publique* (p. 52). Secrétariat général pour la modernisation de l'action publique.

Waller, D. (2020). 10 Steps to Creating a Data-Driven Culture. *Harvard Business Review.*

Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All—A Contingency Approach to Data Governance. *Journal of Data and Information Quality*, *1*(1), 4:1-4:27.

Wende, K., & Otto, B. (2007). *A contingency approach to data governance*.

Wilkinson, N. (2014). *A framework for organisational governance maturity : An internal audit perspective*.

# Détection d'anti-patrons d'alignement dans les SI

## Vers une approche automatisée

**Ali Benjilany [1], Pascal André [1], Hugo Brunelière [2], Dalila Tamzalit [1]**

*[1]Nantes Université, École Centrale Nantes, [2]IMT Atlantique, CNRS, LS2N, UMR 6004, F-44000 Nantes, France*

*Prenom.Nom@ls2n.fr*

RÉSUMÉ. *Les systèmes d'information ont pour rôle de contribuer à l'efficacité des organisations. L'alignement opérationnel des applications avec le métier est un élément clé de la cohérence de ces systèmes. Dans cet article, nous nous intéressons à la détection de potentielles incohérences dans l'alignement entre le métier, plus précisément des processus métier, et les applications qui les implémentent. Nous proposons de détecter de manière automatisée des anti-patrons d'alignement publiés dans la littérature à travers des règles de détection. L'approche est implantée en Archimate dans l'outil Archi et avec le langage de script* `jArchi` *. Nous illustrons la démarche complète d'alignement sur l'exemple* SoftSlate*, un progiciel d'eCommerce. L'ensemble constitue une proposition outillée qui peut être appliquée dans d'autres contextes.*

ABSTRACT. *The role of information systems is to contribute to the efficiency of organisations. The operational alignment of applications with the business is a key element for the coherence of these systems. In this article, we focus on the detection of potential inconsistencies in the alignment between the business, more specifically business processes, and the applications that implement them. We propose to automatically detect alignment anti-patterns published in the literature using detection rules. The approach is implemented in Archimate in the Archi tool and with the* `jArchi` *scripting language. We illustrate the complete alignment process using the example* SoftSlate*, an eCommerce software system. The whole constitutes a tooled proposal which can be applied in other contexts.*

MOTS-CLÉS : *Systèmes d'information - Architecture d'entreprise - Alignement Métier des SI - Anti-Patrons - Métriques*

KEYWORDS: *Information Systems - Enterprise Architecture - Business-IT Alignment - Anti-Patterns - Metrics*

## 1. Introduction

Un enjeu crucial des systèmes d'informations (SI) est qu'ils contribuent à l'efficacité des organisations pour lesquelles ils ont été conçus. Établir l'alignement Business-IT (BITA) d'une organisation est une manière d'évaluer la cohérence entre les métiers (*business*) de l'organisation et le système d'information automatisé (*IT*) support. Cet alignement peut se faire à différents niveaux d'abstraction, des systèmes opérationnels à la stratégie d'entreprise comme expliqué dans un article fondateur (Henderson, Venkatraman, 1999). Pour exprimer ces niveaux d'abstraction, nous nous plaçons dans le cadre de l'architecture d'entreprise (Lankhorst, 2013), appelée aussi urbanisation des SI dans la communauté française (Club, 2010 ; Longépé, 2003). On distingue plusieurs couches selon les modèles d'architecture, par exemple 4 pour le Cigref [1], 5 pour Togaf/Archimate (The Open Group, 2013) et 6 pour Zachman (Zachman, 1987). Quel que soit le modèle, on retrouve une vue métier (*business layer*) et une vue applicative (*application layer*). Ces deux vues sont au centre de l'alignement BITA, et font également partie de l'alignement opérationnel entre le métier et l'IT (Henderson, Venkatraman, 1999). L'alignement opérationnel des applications sur le métier est un élément clé de la cohérence des systèmes d'information. Nous appelons ce périmètre *Core Operational BITA (COBITA)* (André *et al.*, 2023).

Dans ces travaux récents, nous avons mis en évidence l'importance de la représentation des liens inter-couches comme pierre angulaire de l'alignement. L'analyse de cet état des lieux a ainsi mis en évidence l'importance des défis qui restent ouverts et le manque crucial d'outillage pour établir une cartographie de l'alignement COBITA d'une organisation mais également pour évaluer son état. Parmi les approches permettant de diagnostiquer la qualité de l'alignement détaillées dans (André *et al.*, 2023), nous nous focalisons ici sur l'évaluation de l'état de cohérence de l'alignement entre la couche métier et la couche applicative. La cohérence signifie sommairement que l'implantation du SI suit la logique de son organisation métier. La cohérence est définie par un ensemble de propriétés qu'il faut vérifier, par exemple, si deux activités d'un processus métier collaborent alors les éléments applicatifs qui les mettent en œuvre doivent être reliés. Vérifier la cohérence de l'alignement peut s'avérer complexe, surtout lorsque plusieurs points de vue sont pris en compte, tels que les données, les fonctions, les domaines, les performances ou la sécurité. Une autre approche est qu'il est parfois plus aisé de détecter les incohérences que de prouver la cohérence. Concrètement, cela se fait souvent par la mise en évidence de contre-exemples issus de la pratique BITA que de chercher à vérifier la cohérence de l'ensemble de l'état d'alignement. C'est la voie que nous avons choisie dans le travail présenté ici. Pour ce faire, nous considérons les anti-patrons d'alignement métier-IT identifiés dans (Gouigoux, Tamzalit, 2021). Même si ce travail offrent une représentation et une analyse de ces anti-patrons, il manque d'assistance pour les détecter. Nous proposons une approche outillée pour combler ce manque.

1. https://www.cigref.fr/cigref_publications/RapportsContainer/Parus2003/2003_-_Accroitre_l_agilite_du_systeme_d_information_web.pdf

L'article est organisé comme suit. Dans la section 2, nous posons le contexte nécessaire à la compréhension de l'article (modélisation, liens d'alignement, anti-patrons). Ces éléments sont ensuite repris et illustrés sur un exemple, le cas d'étude `SoSl`, basé sur le logiciel `SoftSlate`, dans la section 3. Le cœur de la contribution, à savoir la mise en œuvre des anti-patrons et l'implémentation sont détaillées dans la section 4. Enfin, nous en analysons les résultats d'expérimentations sur le cas d'étude dans la section 5. Nous en discutons la validité dans la section 6 et la comparaison avec d'autres travaux en section 7 avant de conclure.

## 2. Contexte et approche

L'alignement *business/IT* (BITA) est une étape de la démarche d'architecture d'entreprise et d'urbanisation (Aversano *et al*., 2013). Elle permet de cartographier la situation à un instant *t* en visant à obtenir une image fidèle de l'état des sous-ensembles du système d'information. Dans ce contexte, nous nous focalisons sur le cœur de l'alignement, appelé **Core Operational BITA**, CO-BITA en abrégé (André *et al*., 2023). Il s'agit de traiter la relation entre les couches métier et applicative.

Concernant la modélisation des couches, le premier point à fixer est celui du choix du langage de modélisation et le second celui de l'instanciation des modèles. Nous avons déjà identifié que BPMN est le langage majoritaire pour la description des *business processes* de la couche métier tandis que pour la couche application, si UML reste majoritaire, différents types de diagramme sont utilisés (André *et al*., 2023). Pour ce travail, nous avons fait le choix d'utiliser Archimate pour différentes raisons : (i) il s'agit d'un standard pour l'architecture d'entreprise ; (ii) il intègre des langages pour plusieurs couches et des relations génériques entre couches ; (iii) sa couche business est plus épurée que BPMN, (iv) il dispose d'un outil open source de référence `Archi` (même si d'autres outils tels que VisualParadigm ou SmartEA existent sur le marché) ; (v) il dispose d'une extension `jArchi` qui permet d'écrire des scripts. A noter que, pour l'étude de cas traitée dans le papier, les modèles ont été construits manuellement et n'ont pas été automatiquement produits, par exemple par rétro-ingénierie (Aversano *et al*., 2010 ; Pepin *et al*., 2016).

Pour l'évaluation de l'alignement, nous ciblons l'analyse de la cohérence de l'alignement et plus précisément la détection d'incohérences, dans la lignée des travaux visant à définir des **anti-patrons** d'alignement (Gouigoux, Tamzalit, 2021). A l'instar des *anti-patterns* en génie logiciel qui sont des réponses à des erreurs courantes de conception des logiciels (Brown *et al*., 1998), les anti-patrons BITA reprennent des erreurs récurrentes d'alignement. Ces erreurs ont été identifiées sur la base d'études (audit) d'une trentaine de systèmes d'information. Elles ont ensuite été catégorisées sous forme d'anti-patrons. Les auteurs ont identifié 14 anti-patterns BITA dans les systèmes d'information considérés, correspondants chacun à un mauvais scénario récurrent. Ils en présentent quatre et les ont formalisés sous forme de cartes d'identités incluant : Label - Définition - Visualisation - Causes - Conséquences - Effets sur

l'évolution, sur l'utilisation, la facilité et le temps de récupération. Dans les sections suivantes, nous résumons cela aux définitions et visualisations des anti-patterns.

Il manque à cette approche des outils de détection, ce que nous proposons dans la suite de cet article. Mais commençons par présenter le cas d'étude support et sa cartographie des couches métier-applicatif.

## 3. Un exemple d'alignement COBITA : le cas SoSl

Dans cette section, nous présentons le cas d'étude que nous avons construit sur le progiciel `SoftSlate`, qui permet de créer et personnaliser son propre site web e-commerce. Il s'agit d'une application web développée en Java avec le framework J2EE. Elle présente l'avantage de fournir (i) le code source de développement [2]; (ii) un Guide d'utilisateur et un Guide d'administrateur [3]. Notre objectif est de modéliser les couches métier et applicative du système et cartographier les liens entre ces deux couches, avant de chercher à détecter de manière automatisée des situations de mauvais alignements. Dans la suite nous présentons les modèles métier et applicatif de `SoSl`, ainsi que la cartographie représentative de son état d'alignement.

**Modèle métier** L'article (Di Francescomarino *et al.*, 2009) propose une découverte des processus métier par analyse dynamique des applications en suivant les actions des utilisateurs. L'approche est illustrée sur le progiciel `SoftSlate`. C'est donc, *a priori* un point d'entrée intéressant pour nous. Nous n'avons pas eu accès à un modèle métier décrit en BPMN mais simplement aux illustrations de l'article. La couche métier est représentée par le processus métier `Softslate Commerce-recovered BPMN` qui décrit les actions que peut réaliser le client comme Login – Register - SaveCartItem. Néanmoins le modèle proposé est insuffisant. Nous avons donc utilisé la documentation `SoftSlate` pour modéliser deux processus métier : Order annoté `BP1` (Ajouter un produit à la liste des produits listés dans la boutique puis passer la commande de ce produit) et Shipping annoté `BP2` (Configurer les coûts de livraison) avec Archi, comme illustré dans la partie haute de la FIGURE 1 [4]. Notre modélisation reste limitée, il manque notamment la partie réservée au client par manque de documentation relative à cet utilisateur du système.

**Modèle applicatif** Nous avons modélisé la couche applicative en examinant le code source disponible de `SoftSlate`. Le code est composée de cinq packages principaux : Client, Administration, Business Objects, DAO, Installer . Le package Customer et le package Administrator fournissent les fonctionnalités de base du système. Les packages Business Object et DAO sont également importants car ils sont directement utilisés par le package Administrator. Cependant, ces cinq packages contiennent un grand nombre de sous-packages (46 au total), chacun d'entre eux contenant un nombre important de classes. Ainsi, dans le cadre des expérimentations actuelles, nous nous

2. https://www.softslate.com/category/archivedDocs
3. https://www.softslate.com/documentation/userGuide2x.pdf
4. Les deux processus sont détaillés dans la FIGURE 10 et la FIGURE 11 de l'annexe Web[5]

sommes concentrés sur le package Administrator. Il contient un ensemble de classes Java de différents types en suivant le modèle de conception logicielle `Model-View-Controller`. Les classes Servlet implémentent la partie `Controller`, les classes JSP implémentent la partie `View`, tandis que les classes Bean, Processor et DAO implémentent la partie `Model`. De plus, le package Administrator est responsable de l'implémentation des processus métier mentionnés précédemment `BP1` et `BP2`. Le modèle applicatif `ArchiApp` est reporté sur la partie basse de la FIGURE 1. Il est à noter que l'objectif de cette figure est de montrer une vue d'ensemble de l'état d'alignement de notre étude de cas et non de montrer le détail des concepts des deux couches. Ainsi, la représentation des différents liens entre la couche métier et la couche applicative montre une certaine intrication qui reste difficile à analyser manuellement.

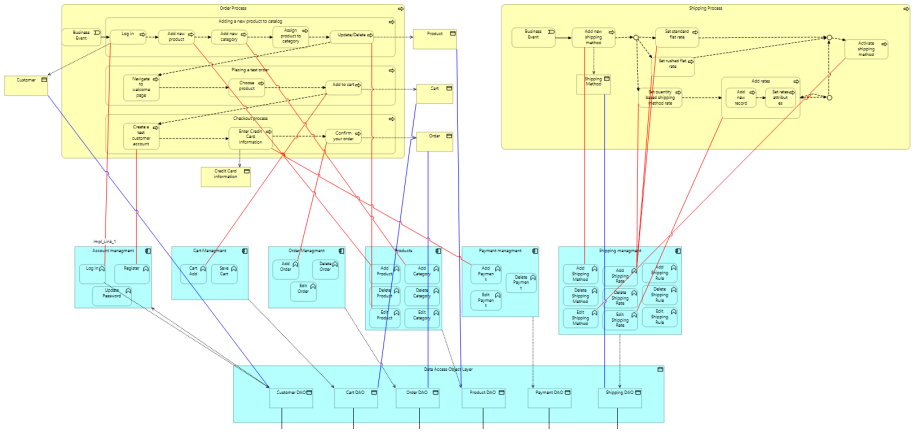


FIGURE 1. *Cartographie d'alignement du cas* `SoS1`.

**Cartographie et état de l'alignement du cas `SoS1`** Les couches 'métier' et 'application' étant modélisées, nous nous pouvons identifier puis modéliser les liens entre elles. La Figure 1, en plus de la couche métier (partie haute) et la couche applicative (partie basse), modélise les liens que nous avons identifiés entre les deux couches en analysant "manuellement" l'application. Ces liens sont de deux types, définis dans (Benjilany *et al.*, 2024) : *implémentation* (en rouge) qui permet de relier un concept de la couche métier à un concept applicatif qui l'implémente dans la couche applicative, et *représentation* (en bleu) qui permet de relier un concept représentant une donnée (exemple: une commande) de la couche métier à un concept représentant une donnée (exemple : un DAO commande) dans la couche applicative. La cartographie obtenue met en évidence une certaine complexité des liens inter-couches. Tenter d'y déceler à l'oeil humain un ou des anti-patrons d'alignement reste un exercice difficile. Nous appellerons cette cartographie le modèle du cas **`SoS1`**. Nous proposons dans la section suivante une approche automatisée de détection de ces anti-patrons.

## 4. Anti-Patrons : mise en œuvre d'une détection automatisée

Les algorithmes sont spécifiés en pseudo-code. Pour chacun des anti-patrons, nous donnons la définition et son identité visuelle issues de (Gouigoux, Tamzalit, 2021), un exemple sur le cas d'étude **`SoSl`** et un algorithme pour le détecter. Les algorithmes proposés sont ensuite implantés avec le langage de script *jArchi* pour être exécutés dans l'environnement de modélisation *Archi*. Ces algorithmes sont accessibles sur l'annexe Web dédiée [5].

### *4.1. Anti-patron AP1: "Pure technical integration"*

DÉFINITION 1 (Pure technical integration). — *L'anti-patron Pure technical integration est présent lorsque les applications logicielles s'appellent les unes les autres directement ou par l'intermédiaire d'autres logiciels, sans que ces liens apparaissent au niveau de la couche métier.*

L'identité visuelle de AP1 est représentée dans la partie gauche de la FIGURE 2. L'existence d'un lien entre concepts applicatifs sans que ce lien n'ait d'équivalent entre les concepts métiers correspondants (en lien avec les concepts applicatifs en question) alerte sur une potentielle erreur d'alignement.

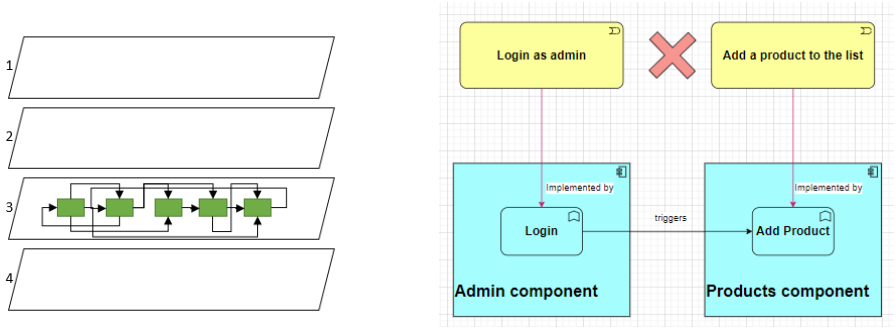



FIGURE 2. *Anti-patron AP1: "Pure technical integration"*

**Illustration concrète.** La partie droite de la FIGURE 2 montre que la fonction applicative Login déclenche la fonction applicative Add product. Ces deux fonctions implantent respectivement deux processus métiers Login as admin et Add a product to the list. Seulement, le lien de déclenchement existant entre les deux fonctions applicatives n'est pas retranscrit entre les deux processus métiers correspondants.

**Algorithme de détection.** Dans Listing 1 ci-dessous, la première étape (STEP 1) crée une collection des concepts alignables de la couche métier (notée bfuncConcepts). La deuxième étape (STEP 2) vérifie que chacun de ces concepts est lié directement à un concept de la couche applicative (variable imp. target). La troisième étape (STEP 3)

5. https://uncloud.univ-nantes.fr/index.php/s/5ZrjSnbdb3neBTP

filtre les concepts applicatifs (variable rel . target ) liés au concept applicatif imp. target par des liens de déclenchement. La quatrième étape (STEP 4) cherche les concepts métiers (variable impl. source) reliés à chaque concept applicatif filtré dans STEP 3. La dernière étape (STEP 5) vérifie si chaque concept métier issu de STEP 4 est exactement le concept métier en cours de traitement dans STEP 2. Les liens manquants sont une source potentielle d'anti-patron AP1.

Listing 1 – Pseudo-code pour Anti-Patron 1

```
// STEP 1 : Collect all B func concepts
busprocess = GetElement("business-process")
busfunction = GetElement("business-function")
businteraction = GetElement("business-interaction")
bfuncConcepts = Concat(busprocess, Concat(busfunction,
    businteraction))
collection = GetElement(bfuncConcepts)
ForEach concept in collection
 // STEP2: each BP is implemented by at least one application
 ForEach imp in GetRels(concept, 'association-relationship')
  If imp.specialization is 'Implementation link'
  // STEP3: Look for the links between AF1 and AF2, AF i ?
   ForEach rel in GetOutRels(imp.target, 'triggering-
       relationship')
    writeln(imp.target, 'is related by', rel, 'to', rel.target)
    // STEP4: For each AFi: Retrieve the BPi implemented.
    ForEach impl in GetRels(rel.target, 'association-
        relationship')
     If imp.specialization is 'Implementation link'
      writeln('which implements', impl.source)
      // STEP5: Check if BPi is related to BP1
      ForEach flow in GetRels(impl.source, 'flow-relationship')
       If flow.source.id not = concept.id
        writeln('WARNING AP1 detected! There should be a flow
            link FROM', concept, 'TO', flow.source)
       End If
      End ForEach
     End If
    End ForEach
   End ForEach
  End If
 End ForEach
End ForEach
```

### *4.2. Anti-patron AP2: "The functional silo dedicated IT subsystem"*

DÉFINITION 2. — *L'anti-patron "functional SILO" se réfère à l'existence d'un bloc isolé du reste du SI. Un bloc est caractérisé par une succession de tâches métiers reliée à une succession de fonctions applicatives.*

L'identité visuelle d'AP2 est représentée dans la partie gauche de la FIGURE 3. Une partie du système d'information automatisé est complètement isolée du reste. Cela peut être volontaire, comme dans certains contextes réglementaires stricts, mais cela peut également conduire à la duplication de fonctions dans d'autres parties du système, voir à des parties jamais utilisées.

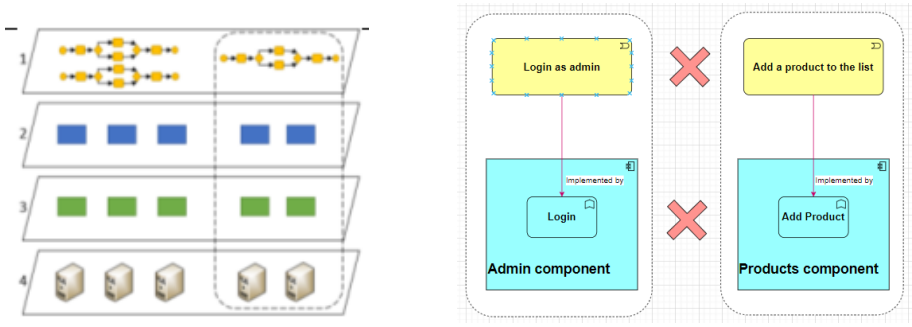



FIGURE 3. *Anti-patron AP2: "The functional SILO dedicated IT subsystem"*

**Illustration concrète.** La partie droite de la Figure 3 montre deux blocs isolés. Le constat est que ces deux blocs n'ont respectivement aucun lien entre eux. En itérant l'observation pour chaque bloc sur tout le système et, qu'au final, au moins un des blocs est isolé, alors nous sommes face à une illustration de l'anti-patron AP2.

**Algorithme de détection.** L'algorithme pour détecter AP2 (Listing 3 de l'annexe Web[5]) commence par créer une collection bfuncConcepts des concepts de la couche métier. On vérifie ensuite pour chaque concept métier l'absence de lien (flux ou composition) vers d'autres concepts métiers (implémentation directe ou indirecte). En cas d'absence, si le concept est implanté (présence de liens d'implémentation) par des concepts applicatifs cibles de ces liens et reliés à d'autres concepts applicatifs de type application −function alors ces derniers ne doivent pas être l'implantation d'autres concepts métiers sinon une alerte est levée pour indiquer la présence de l'anti-patron 2 sur le concept métier considéré.

### *4.3. Anti-patron AP3: "Monolith application"*

DÉFINITION 3. — *Une application monolithique se réfère à un seul et même concept applicatif qui implémente plusieurs fonctions métiers distinctes.*

L'identité visuelle pour l'anti-patron 3 est représentée dans la partie gauche de la Figure 4. L'existence d'au moins deux liens partant vers un seul et même concept applicatif alerte sur une éventuelle erreur d'alignement.

**Illustration concrète.** La partie droite de la FIGURE 4 montre que le concept applicatif Login implémente deux concepts métiers distincts Login as admin et Login as customer. Les deux concepts métier d'authentification sont implémentés dans le même composant applicatif, que ce soit un administrateur ou un client qui sollicite ce service.

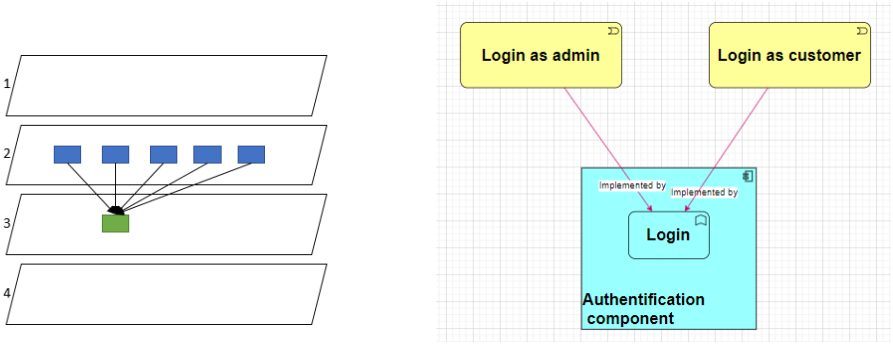



FIGURE 4. *Anti-patron AP3: "Monolith application"*

**Algorithme de détection.** L'algorithme pour identifier AP3 (Listing 4 de l'annexe Web[5]) crée une collection de tous les concepts que nous considérons dans la couche applicative. Pour chaque concept, on filtre, dans la collection des liens d'association, les liens d'implantation directs ou indirects vers la couche métier. S'il n'y en a pas c'est un problème, s'il y en a plusieurs c'est un potentiel cas d'anti-patron AP3.

### *4.4. Anti-patron AP4: "Functional multiple implementations"*

DÉFINITION 4. — *L'implémentation fonctionnelle multiple se réfère à l'implantation d'un seul concept de la couche métier par plusieurs concepts de la couche applicative. La principale cause évoquée est le manque de compréhension des modèles métiers (partie fonctionnelle en particulier).*

L'identité visuelle d'AP4 est montrée dans la partie gauche de la Figure 5. L'existence de deux ou plusieurs liens partant d'un seul et même concept métier alerte sur une potentielle erreur d'alignement.

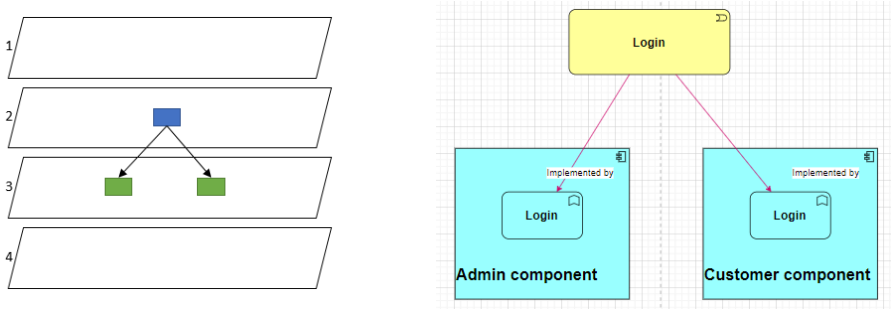



FIGURE 5. *Anti-patron AP4: "The multiple functional implementation"*

**Illustration concrète** La partie droite de la Figure 5 montre que le concept métier Login est implémenté par deux concepts applicatifs Login du composant applicatif Admin et Login du composant applicatif Customer. Dans le travail initial, l'exemple

donné est celui du concept métier reporting qui est implémenté de 40 façons différentes par les applications concernées. Dans une cartographie manuelle où plusieurs autres liens sortant/entrant du/au même concept sont représentés, repérer visuellement une telle situation nécessite beaucoup d'efforts, est de surcroît aléatoire et peut même s'avérer impossible dans certains cas.

**Algorithme de détection** : L'algorithme pour identifier AP4 ((Listing 4 de l'annexe Web[5]) est composé de deux étapes. La première crée une collection de tous les concepts de la couche métier à traiter. Pour chaque concept on détecte les liens manquants vers la couche applicative, ou les liens multiples d'implantation, source potentielle d'un anti-patron AP4.

## 5. Expérimentation et Analyse des résultats

Dans cette section, nous présentons les expérimentations réalisées sur le cas d'étude **SoSl** avec les algorithmes décrits précédemment. Dans un premier temps nous commentons et analysons les premiers résultats obtenus par l'outil de détection qui implante les algorithmes en `jArchi`. Dans un second temps, nous améliorons la confiance dans les algorithmes par une analyse de mutation.

**Application au cas SoSl, commentaires et premières analyses** L'exécution des scripts lève une alerte seulement pour AP3, *cf.* FIGURE 6. Une alerte "WARNING" est levée car le concept application Add Shipping Rate possède 3 liens d'implémentation vers 3 concepts métiers différents. Les autres situations ne détectant pas AP3 sont ignorées.

```
 The A-func-Concept :  Add Shipping Method  has : 1 implementation links

 The A-func-Concept :  Edit Shipping Method  has : 1 implementation links
The A-FUNC-CONCEPT Delete Shipping Method IS NOT IMPLEMENTING ANY BFC

 The A-func-Concept :  Add Shipping Rate  has : 3 implementation links
WARNING Monolith application BITA anti-pattern
```

FIGURE 6. *Détection d'AP3 sur* `SoSl`

Ces résultats peuvent être interprétés de plusieurs façons. Une première façon consiste à en tirer des indications sur le niveau d'alignement du système : l'existence d'une seule occurrence de lien d'implémentation peut être interprétée comme indicateur d'un relativement bon alignement du système. Cependant, il est difficile d'avancer un taux d'alignement réaliste sur cette seule base. Par exemple, nous proposons déjà par ailleurs d'évaluer de manière complémentaire le niveau d'alignement sur la base d'un ensemble de métriques et de règles de cohérence (Benjilany *et al.*, 2024).

Deux aspects peuvent également impacter le résultat de nos expérimentations: (i) *Qualité de la modélisation.* Le modèle `SoSl` est incomplet (deux processus métiers seulement) et non validé (notre modèle n'est pas contrôlé par des équipes métiers

et des équipes de développement). De plus, nous ne sommes pas à l'abri d'erreurs d'interprétation de notre part. (ii) *Qualité des algorithmes de détection.* Les algorithmes de détection que nous proposons peuvent ne pas détecter des erreurs d'alignement présentes dans le système (faux négatifs) ou détecter des erreurs qui, en réalité, n'en sont pas (faux positifs). Pour pallier le potentiel manque d'efficacité des algorithmes, nous nous inspirons des techniques de mutation dans le test logiciel pour créer différentes situations de mauvais alignement permettant de tester les algorithmes.

**Validation par mutation** Pour vérifier si les anti-patterns sont détectés, nous avons donc créé des mutants du modèle initial `SoSl` qui contiennent des erreurs d'alignement. Pour tester chaque algorithme de détection d'anti-patron (AP), nous avons défini deux mutants : un premier mutant `A` où nous injectons volontairement une simple occurrence de l'AP ciblé (v1) et un deuxième mutant `B` où nous injectons volontairement soit une occurrence plus complexe soit plusieurs occurrences de l'AP ciblé (v2). Le TABLEAU 1 présente les mutants et les résultats de ces tests. Ces mutants, ainsi que le modèle `SoSl` de base, sont tous accessibles depuis l'annexe Web[5]. Les quatre dernières colonnes rapportent le nombre d'erreurs relevées pour chaque exécution des scripts de détection d'AP sur chacun des mutants. Il est à noter que l'AP3 est détecté sur tous les mutants par héritage du modèle de base ; il n'y a donc pas eu de régression. Les résultats montrent que toutes les occurrences d'anti-patrons injectées ont été détectées et qu'aucun faux négatif n'est apparu.

TABLEAU 1. *Démarche des expérimentations et résultats*

| **Base / Mutant** | **Cible Variante** | **Modification(s) effectuées sur le modèle de base** | **Résultats** | | | |
|---|---|---|---|---|---|---|
| | | | AP1 | AP2 | AP3 | AP4 |
| `SoSl` | base héritée | Aucune | 0 | 0 | 1 | 0 |
| `SoSl 1A` | AP1 v1 | Supprimer le lien flux entre Log In et Add Product | 1 | 0 | 1 | 0 |
| `SoSl 1B` | AP1 v2 | Injecter un nouveau BP LIS | 2 | 0 | 1 | 0 |
| `SoSl 2A` | AP2 v1 | Injecter et Isoler BP LIS implémenté par AF isolé | 0 | 1 | 1 | 0 |
| `SoSl 2B` | AP2 v2 | Injecter et isoler "Isolated BP" implémenté par "Isolated AF" | 0 | 2 | 1 | 0 |
| `SoSl 3A` | AP3 v1 | AF Login implémente 2 BP | 0 | 0 | 2 | 0 |
| `SoSl 3B` | AP3 v2 | AF Login implémente 4 BP | 0 | 0 | 2 | 0 |
| `SoSl 4A` | AP4 v1 | BP Login implémenté par 2 AF | 0 | 0 | 1 | 1 |
| `SoSl 4B` | AP4 v2 | BP Login implémenté par 5 AF | 0 | 0 | 1 | 1 |

Illustrons maintenant sur des exemple pour AP1, AP2 et AP4.

La FIGURE 7 montre le résultat de l'exécution du script pour identifier AP1 sur le mutant représenté en partie droite de la FIGURE 2. Un message apparaît pour annoncer qu'un lien (flux) est manquant entre les deux concepts métiers Login as admin et Add new product.

La Figure 8 montre le résultat de l'exécution du script pour identifier AP2 sur le mutant représenté en partie droite de la FIGURE 3. Nous pouvons y voir une alerte "WARNING" notifiant d'une éventuelle erreur d'alignement relative au concept métier Login as seller . Les autres concepts métiers, en revanche, ne sont pas concernés.

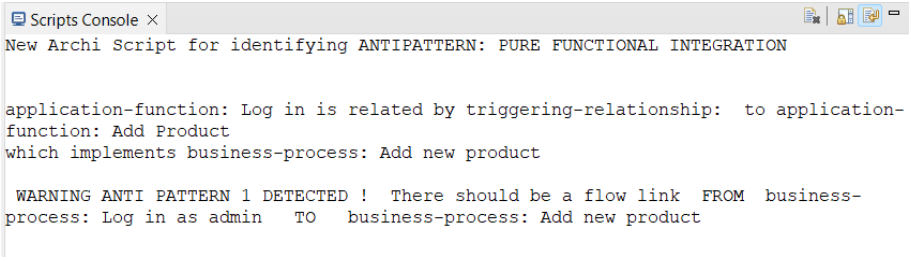



FIGURE 7. *Détection d'AP1 sur* `SoS1 1A`

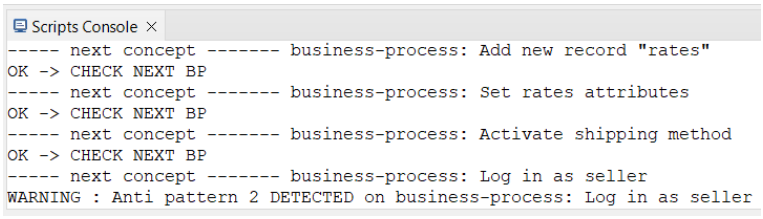



FIGURE 8. *Détection d'AP2 sur* `SoS1 2A`

La figure 9 montre le résultat de l'exécution du script pour identifier AP4 sur le mutant représenté en partie droite de la FIGURE 5. Nous pouvons y voir qu'une alerte "WARNING" indiquant que le concept métier Login est implémenté par deux liens.

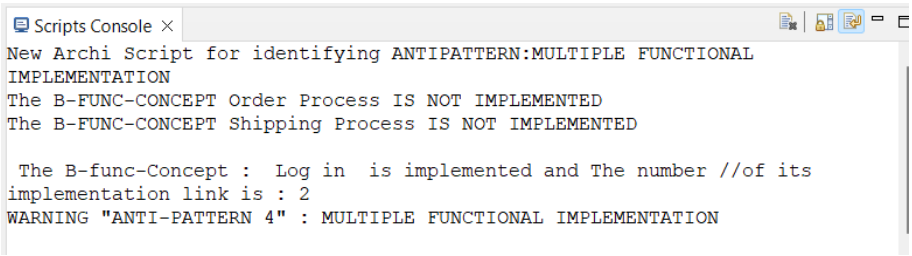



FIGURE 9. *Détection d'AP4 sur* `SoS1 4A`

## 6. Discussion

Dans cette section, nous évaluons plus en détail les résultats obtenus lors des expérimentations présentées dans la Section 5 afin d'identifier les limites actuelles de notre proposition. Sur cette base, nous indiquons plusieurs pistes possibles d'amélioration.

**Évaluation et limitations** Pour être applicable, notre proposition requiert l'existence de modèles des couches métier et applicative. Cependant, de tels modèles ne sont pas toujours disponibles (notamment les modèles métiers) et doivent souvent être construits à partir de différentes sources d'information. Pour notre cas d'étude, nous disposions uniquement de la documentation utilisateur et du code source de l'application. A partir de ces éléments, nous avons construit manuellement les modèles métier

et applicatif en utilisant le langage de modélisation Archimate. Une telle approche s'avère coûteuse en temps dans le cadre d'un système d'information de taille plus conséquente. Ainsi, les expérimentations réalisées jusqu'à présent montrent l'applicabilité de notre proposition. En revanche, son passage à l'échelle reste encore à valider. Les résultats obtenus démontrent également que notre proposition permet de détecter les occurrences potentielles des quatre AP supportés dans le contexte d'un cas d'étude comme **`SoSl`**. Cependant, la seule information concernant la possible présence ou non d'AP n'est pas suffisante pour pouvoir statuer sur l'état d'alignement du système étudié. Les anti-patrons ne constituent qu'un outil parmi d'autres permettant de déterminer le niveau d'alignement d'un système (André *et al*., 2023). En effet, ceux-ci doivent être complétés par l'utilisation de métriques et/ou de règles de cohérence architecturale (par exemple) afin de pouvoir calculer un véritable taux d'alignement. Enfin, nous n'avons traité qu'un cas d'étude jusqu'à présent. D'autres expérimentations seront nécessaires pour augmenter la confiance en nos algorithmes.

**Améliorations possibles** Une première piste d'amélioration consiste à élargir le spectre des langages de modélisation pouvant être utilisés dans notre proposition. Ainsi, en complément d'Archimate, d'autres standards de modélisation tels que BPMN et UML pourraient être intégrés pour supporter les couches métier et applicative respectivement. Cependant, cela nécessite de disposer d'un environnement de modélisation en mesure de permettre l'utilisation conjointe de ces différents langages. Ce n'est pas le cas de l'outil de modélisation Archi sur lequel repose actuellement notre proposition. Une seconde piste d'amélioration consiste à perfectionner les algorithmes de détection des anti-patrons. Nous fournissons pour le moment quatre algorithmes supportant quatre AP distincts. Les premières expérimentations ont permis de s'assurer de leur fonctionnement individuel. Cependant, leur complexité actuelle pourrait être réduite afin de permettre leur utilisation dans le contexte de cas de taille plus conséquente. Un travail d'optimisation des algorithmes (e.g., concernant le nombre de boucles imbriquées) est donc à prévoir lors de nos prochaines étapes. En parallèle, la combinaison de plusieurs algorithmes peut être envisagée pour des raison de performance mais aussi dans l'optique de l'identification d'AP possiblement interdépendants. Enfin, ces algorithmes pourraient être revus afin d'enrichir leurs résultats, e.g., pour fournir des recommandations d'évolution concernant l'une ou l'autre des couches concernées. Une troisième piste d'amélioration consiste à compléter la librairie d'outils à disposition pour évaluer l'état de l'alignement. Des outils complémentaires, tels que des métriques ou encore des règles de cohérence, pourraient être utilisés. Là encore, leur combinaison avec les algorithmes déjà disponibles devrait être travaillée afin de permettre le calcul d'un taux d'alignement le plus représentatif possible de la réalité.

## 7. Travaux connexes

En partant de la revue de littérature la plus récente sur le sujet du *mis-alignment* (Őri, Szabó, 2024), nous positionnons notre approche suivant deux axes : (i) le niveau d'outillage disponible et (ii) les techniques de détection utilisées.

Selon cette revue, six travaux proposés rentrent dans le périmètre de l'alignement opérationnel COBITA. Parmi ceux-ci, quatre s'intéressent spécifiquement à la détection du non-alignement. Une thèse évoque le non-alignement entre des objectifs stratégiques et opérationnels (Singh, 2009). Bien que partiellement outillée, l'approche proposée se différencie de notre proposition par son positionnement au niveau de la couche stratégique et non des processus métiers. Un autre travail présente les résultats d'une investigation réalisée auprès d'employés de 7 départements d'une entreprise (Peng *et al*., 2021). Sur la base de cette investigation, un modèle de non-alignement est proposé. Cependant, contrairement à notre proposition, ce modèle n'est pas accompagné d'un support outillé permettant sa mise en œuvre. De manière similaire, une étude recense des pratiques de gestion visant à atténuer les facteurs de risque de non-alignement (Mamoghli *et al*., 2015). Cette fois encore, les pratiques identifiées ne sont pas supportées par de l'outillage technique associé. Une dernière approche propose des recommandations d'amélioration, en plus de la détection (Chen *et al*., 2005). Elle repose sur un protocole en 12 étapes : Les étapes 1 à 9 décrivent comment déterminer une situation de non-alignement, les étapes 10 à 12 permettent de proposer des scénarii de ré-alignement. La démarche est intéressante et les étapes 1 à 9 partagent des objectifs commun avec notre proposition. En revanche, les étapes 10 à 12 sont hors du contexte de notre proposition actuelle. De plus, l'approche globale proposée n'est une nouvelle fois pas outillée.

D'autres approches proposent également différentes techniques de détection de patrons. Par exemple, une approche propose de traiter le problème de l'alignement stratégique via un algorithme pour la détection, la correction et la prévention des situations de non-alignement dans des modèles d'Architecture d'Entreprise (Aseeva *et al*., 2022). L'algorithme proposé s'inspire de la méthodologie Architecture Development Method (ADM) du standard TOGAF, prenant en compte des éléments de niveau stratégique. Notre proposition se différencie donc de cette approche de par son accent mis sur les couches métier et applicative. De plus, l'algorithme proposé n'est pas encore implémenté contrairement à ceux que nous proposons. Un autre travail, s'appuyant sur l'analyse manuelle de plusieurs cas d'étude, propose un *framework* de quatres patrons de conception dans un contexte d'alignment Business-IT (Cleven, 2011). Mais l'objectif principal de ce *framework* est l'évaluation de la performance des processus, et non de l'alignement comme dans notre proposition. De plus, là encore, la solution proposée reste théorique et aucune implémentation technique n'est fournie. Dans la même veine, un travail complémentaire s'intéresse également au problème de la performance dans un contexte d'applications d'Entreprise (Wert *et al*., 2014). Ainsi, cinq heuristiques sont proposées afin de détecter, sur la base de mesures, des anti-patrons de performance connus. Ces heuristiques sont outillées par le biais d'un *framework* technique appelé Dymanic Spotter. Cependant, l'accent est mis encore une fois sur l'aspect performance et non sur l'évaluation plus générale de la qualité de l'alignement. D'un point de vue Génie Logiciel, la détection d'anti-patrons a aussi été étudiée dans le contexte de processus de développement logiciel (Pícha *et al*., 2022). L'outil nommé Software Process Anti-patterns Detector (SPADe) permet une telle vérification sur la base de descriptions semi-formelles d'anti-patrons fournies grâce à un *template*. Ce-

pendant, contrairement à notre proposition, les anti-patrons ainsi définis et détectés ne concernent que les couches basses (e.g., applicative, technologique) sans prendre en compte la couche métier.

## 8. Conclusion

Dans ce papier, nous nous sommes focalisés sur l'évaluation de l'état de cohérence de l'alignement entre la couche métier et la couche applicative, aussi appelé Core Operational BITA (COBITA). Sur la base d'anti-patrons d'alignement identifiés dans des travaux précédents, nous avons proposé un catalogue initial d'algorithmes permettant la détection automatisée d'un certain nombre de ces anti-patrons. Nous avons pu expérimenter en pratique avec l'implémentation de ces algorithmes dans le cadre d'un cas d'étude concret que nous avons au préalable modélisé.

Ces travaux constituent une étape vers le support automatisé pour une évaluation plus globale du COBITA. En effet, le catalogue d'algorithmes déjà fourni pourrait être complété afin de couvrir un ensemble plus large d'anti-patrons. De plus, l'approche proposée pourrait être étendue afin de supporter d'autres langages standards de modélisation, en complément d'Archimate. Enfin, un effort reste encore à faire afin d'être en mesure de combiner ces algorithmes de détection d'anti-patrons avec d'autres techniques telles que des métriques ou des règles de cohérence. L'objectif final est de permettre le calcul d'un taux global d'alignement le plus réaliste possible.

## Bibliographie

André P., Tamzalit D., Benjilany A., Bruneliere H. (2023). A review of core operational business-it alignment. In *ISD 2023 proceedings*, p. 1–12. Lisbon, Portugal, AIS eLibrary.

Aseeva N., Babkin E., Malyzhenkov P., Masi M. (2022). Strategic integration of alignment models for the it-business misalignment detection and redress. In *Digitalization of society, economics and management: A digital strategy based on post-pandemic developments*, p. 97–113. , Springer.

Aversano L., Grasso C., Tortorella M. (2010). Measuring the alignment between business processes and software systems: A case study. In *SAC '10*, p. 2330–2336. , ACM.

Aversano L., Grasso C., Tortorella M. (2013). A literature review of Business/IT alignment strategies. In *Enterprise information systems*, p. 471–488. , Springer.

Benjilany A., André P., Tamzalit D., Bruneliere H. (2024, avril). Towards a link mapping and evaluation approach for Core Operational Business-IT Alignment. In *26th International Conference on Enterprise Information Systems (ICEIS 2024).* Angers, France, Insticc.

Brown W. H., Malveau R. C., McCormick H. W. S., Mowbray T. J. (1998). *Antipatterns: Refactoring software, architectures, and projects in crisis* (1st éd.). USA, John Wiley & Sons, Inc.

Chen H.-M., Kazman R., Garg A. (2005). Bitam: An engineering-principled method for managing misalignments between business and it architectures. *Science of Computer Programming*, vol. 57, n° 1, p. 5–26.

Cleven A. (2011). Exploring patterns of business-it alignment for the purpose of process performance measurement. In *European conference on information systems.* , . Consulté sur https://api.semanticscholar.org/CorpusID:14754780

Club U.-E. (2010). *Urbanisme des si et gouvernance: Bonnes pratiques de l'architecture d'entreprise*. , Dunod.

Di Francescomarino C., Marchetto A., Tonella P. (2009). Reverse engineering of business processes exposed as web applications. In *2009 13th european conference on software maintenance and reengineering*, p. 139–148. , .

Gouigoux J., Tamzalit D. (2021). Business-it alignment anti-patterns: A thought from an empirical point of view. In E. Insfrán *et al.* (Eds.), *Information systems (ISD2021 proceedings).* Valencia, Spain, Universitat Politècnica de València / Association for Information Systems.

Henderson J. C., Venkatraman H. (1999). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems*, vol. 38, n° 2.3, p. 472–484.

Lankhorst M. M. (2013). *Enterprise architecture at work - modelling, communication and analysis (3. ed.).* , Springer.

Longépé C. (2003). *The enterprise architecture it project: the urbanisation paradigm.* , Elsevier.

Mamoghli S., Goepp V., Botta-Genoulaz V. (2015). An operational "risk factor driven" approach for the mitigation and monitoring of the "misalignment risk" in enterprise resource planning projects. *Computers in Industry*, vol. 70, p. 1–12.

Őri D., Szabó Z. (2024). A systematic literature review on business-it misalignment research. *Information Systems and e-Business Management*, p. 1–31.

Peng G., Chen S., Chen X., Liu C. (2021). An investigation to the industry 4.0 readiness of manufacturing enterprises: The ongoing problems of information systems strategic misalignment. *Journal of Global Information Management (JGIM)*, vol. 29, n° 6, p. 1–20.

Pepin J., André P., Attiogbé J. C., Breton E. (2016). An improved model facet method to support EA alignment. *Complex Systems Informatics and Modeling Quarterly*, vol. 9, p. 1–27.

Pícha P., Hönel S., Brada P., Ericsson M., Löwe W., Wingkvist A. *et al.* (2022). Process anti-pattern detection–a case study. In *Proceedings of the 27th european conference on pattern languages of programs*, p. 1–18. , .

Singh S. N. (2009). *A goal-based requirements gathering approach to detect and understand business-it misalignments*. Thèse de doctorat non publiée, University of British Columbia.

The Open Group. (2013). *Archimate 2.1 specification.* , Van Haren Pub.

Wert A., Oehler M., Heger C., Farahbod R. (2014). Automatic detection of performance anti-patterns in inter-component communications. In *International acm sigsoft conference on quality of software architectures.* , . Consulté sur https://api.semanticscholar.org/CorpusID:2212843

Zachman J. A. (1987). A framework for information systems architecture. *IBM systems journal*, vol. 26, n° 3, p. 276–292.

# Écosystème d'Affaires Numériques: Modèle Organisationnel, Rôles et Gouvernance pour la Flexibilité

**Elena KORNYSHOVA** [1], **Laurent BOUTAL** [2], **Mustapha Kamal BENRAMDANE** [1]

*1. CEDRIC, CNAM, 292 rue Saint Martin, 75003 Paris, France*
*elena.kornyshova@cnam.fr,mustapha-kamal.benramdane@lecnam.net*

*2. RATP, 54 Quai de la Rapée, 75012, Paris, France*
*laurent.boutal@ratp.fr*

*RÉFÉRENCE DE L'ARTICLE INTERNATIONAL. Cet article est un résumé de l'article: Kornyshova E., Boutal L., Benramdane M. K. (2023). Digital Business Ecosystems: Organizational Model, Roles, and Governance towards Flexibility, KES 2023, pp. 4621-4630.*

## 1. Introduction

Les Écosystèmes d'Affaires Numériques (en anglais: Digital Business Ecosystems (DBE)) facilitent la collaboration entre les entreprises sur des plateformes numériques, modifiant ainsi les relations B2B traditionnelles. Cependant, cette nouvelle forme d'organisation soulève des défis en matière de gouvernance, notamment en ce qui concerne la gestion des relations et des échanges de données au sein des DBE. Une question clé est comment garantir la flexibilité de ces relations lorsque le partage de données étend les rôles traditionnels de "fournisseur" et de "client" dans la chaîne d'approvisionnement. Pour répondre, nous avons mené une étude et collecté des données auprès du secteur bancaire française pour élaborer un modèle organisationnel des DBE, détaillant les rôles et les règles de coopération. Nous proposons un modèle de haut niveau des DBE, conceptualisons les participants comme des acteurs exerçant simultanément les rôles de client et de fournisseur, et proposons une matrice de gouvernance des DBE (Kornyshova *et al*., 2023). Nous pensons que ce modèle favorisera une distribution équitable des ressources, ainsi que le respect des objectifs communs et individuels au sein des DBE.

## 2. Matrice de Gouvernance

La matrice de gouvernance permet de qualifier les relations entre un DBE et ses acteurs, en définissant les rôles à ces deux niveaux (voir Tableau 1). Ces rôles sont

TABLE 1 – Matrice de Gouvernance des DBE.

| Niveau | Objectifs | Règles | Données | Processus | Produits / Services |
|---|---|---|---|---|---|
| Niveau DBE | Le DBE a des objectifs et aligne ceux des acteurs | Le DBE établit les règles, les diffuse et vérifie leur respect | Le DBE synchronise et intègre les données | Le DBE synchronise et intègre les processus | Le DBE garantit que les produits et services sont disponibles pour tous les acteurs et qu'ils sont conformes |
| Niveau acteurs des DBE | L'acteur a des objectifs et contribue aux objectifs du DBE | L'acteur respecte les règles établies pour le DBE | L'acteur gère/ stocke/ manipule/ partage ses propres données | L'acteur se concentre sur les activités lui apportant de la marge | L'acteur est un producteur et un consommateur de produits et de services |

détaillés en ce qui concerne les fonctions de gouvernance (objectifs et règles) et les couches de l'espace numérique (données, processus, produits/services). Un DBE est considéré comme un système d'acteurs DBE ayant des objectifs conformes aux objectifs du DBE. Cela permet un espace multicouche facilitant l'échange mutuel de produits/services, de processus et de différentes ressources, comme les données. Les relations entre clients et fournisseurs dans les DBE sont récursives, chaque acteur pouvant être à la fois client et fournisseur. Un tel schéma instaure une compétition et une coopération entre les acteurs du DBE en co-créant et en co-consommant des produits, des services, des processus, et des ressources. La gestion des données soulève des questions complexes liées à la propriété, à la confidentialité et à la sécurité. La gouvernance doit donc établir des normes et des protocoles clairs pour garantir la protection et la confidentialité des données tout en facilitant leur échange et leur utilisation efficaces. Elle doit également aborder les questions de conformité réglementaire et de responsabilité juridique, notamment en mettant en place des mécanismes de conformité robustes pour éviter les sanctions juridiques. Du fait que les acteurs soient différents en termes de taille et de secteur de marché, la gouvernance doit favoriser un accès équitable aux ressources et promouvoir la collaboration entre les acteurs.

## 3. Conclusion

Nous proposons une organisation des DBE en deux niveaux. La gouvernance dans les DBE facilite la collaboration et la création de valeur entre les acteurs. Elle repose sur des normes et des processus clairs qui garantissent la transparence, la responsabilité et la conformité réglementaire, œuvrant ainsi pour la durabilité du DBE.

## Bibliographie

Kornyshova E., Boutal L., Benramdane M. K. (2023). Digital business ecosystems: Organizational model, roles, and governance towards flexibility. *27th International Conference on Knowledge Based and Intelligent Information and Engineering Sytems (KES 2023)*, vol. 225, p. 4621-4630.

# Interopérabilité des métadonnées de la Science Ouverte : qu'en est-il de la réalité ?

Vincent-Nam Dang[1], Nathalie Aussenac-Gilles[2], Imen Megdiche[3], Franck Ravat[1]

*1. IRIT, CNRS (UMR 5505), Université Toulouse Capitole, France*
*vincent-nam.dang@irit.fr, franck.ravat@irit.fr*
*2. IRIT, CNRS (UMR 5505), France*
*nathalie.aussenac-gilles@irit.fr*
*3. IRIT, CNRS (UMR 5505),* INU Champollion, ISIS Castres, Université de Toulouse, France
imen.megdiche@univ-jfc.fr

*REFERENCE DE L'ARTICLE INTERNATIONAL Cet article est un résumé de l'article : Dang, V. N., Aussenac-Gilles, N., Megdiche, I., & Ravat, F. (2023, May). Interoperability of Open Science Metadata: What About the Reality?. In International Conference on Research Challenges in Information Science (pp. 467-482). Cham: Springer Nature Switzerland.*

La Science Ouverte est un mouvement de la recherche scientifique visant à mettre en place des échanges d'informations inter-communautaires et inter-domaines. Ces échanges ont pour objectif d'enrichir le processus de création de connaissance de la recherche. Ce mouvement promeut en particulier la mise à disposition auprès de tous des jeux de données, algorithmes, codes informatiques et processus méthodologiques utilisés et produits par les recherches présentées dans des articles. Pour en faciliter la recherche et la réutilisation, il est prévu que ces jeux de données soient décrits par des métadonnées et rendus accessibles via des plateformes de partage sur le Web. Cependant, l'utilisation conjointe de plusieurs jeux de données venant de différentes sources suppose qu'ils sont interopérables. La grande variété de modèles de métadonnées utilisés dans la Science Ouverte ne permet pas une interopérabilité native de ces métadonnées. Une solution est apportée par les outils de mappings automatiques de modèles de métadonnées. Cependant, on observe un écart entre les performances observées de ces outils dans les campagnes d'évaluations et leur utilisation sur des données réelles. Pour comprendre précisément les besoins liés à la mise en place de la Science Ouverte, nous avons exploré l'état des lieux de l'interopérabilité des métadonnées dans la Science Ouverte.

Dans un premier temps, nous avons mené une étude comparative des différentes définitions en nous basant sur les critères relatifs à une théorie formelle de l'interopérabilité. Aucune des définitions ne permet de répondre à tous les critères.

Nous avons donc proposé une définition de l'interopérabilité répondant à ces critères et permettant de répondre aux problématiques soulevées par les autres définitions. Nous définissons l'interopérabilité comme « la capacité de 2 entités communicantes à travailler en collaboration à travers un échange d'informations pour réaliser un objectif ». Cette définition repose sur 3 aspects : le contexte de la communication, les entités communicantes et l'objectif à réaliser. Nous avons proposé un modèle en 7 couches, nécessaires à la mise en place d'une interopérabilité. Chaque couche possède un objectif précis ainsi qu'une liste de mécanismes d'interopérabilité associés, dépendants du contexte, des entités communicantes et de l'objectif de la communication. On observe 2 catégories de mécanismes d'interopérabilité : les mécanismes de standardisation et les mécanismes de mise en place de passerelles.

Nous avons exploré la dernière couche de ce modèle, la couche de croisement (mise en correspondance ou mapping). L'objectif de cette couche est la mise en place d'un croisement des informations entre les 2 entités communicantes. Dans le contexte de la Science Ouverte, il s'agit de mettre en place un croisement des modèles de métadonnées pour permettre une recherche sur plusieurs plateformes de gestion de données. Nous avons évalué plusieurs outils, comme COMA provenant de la campagne d'évaluation contrôlée OAEI, afin de voir si les outils de mappings automatiques sont capables de mettre en place une interopérabilité automatique sur des métadonnées provenant de plateformes de gestion de données de la Science Ouverte. Les expérimentations ont montré leur inadéquation au monde de la Science Ouverte, avec un F1-score maximum de 0.21, malgré les bons résultats observés dans leur évaluation initiale. Nous avons réalisé une étude statistique de ces résultats. Il ressort une corrélation négative entre la taille du modèle et les performances des outils. De plus, nous avons observé que seules les métadonnées générales et techniques ont réussi à être correctement appariées.

Nous en avons retenu 3 pistes potentielles pour une amélioration de l'intéropérabilité des métadonnées : (i) la classification des métadonnées (générales, techniques et spécifiques à un domaine) permettant une sélection de sous-modèles pour effectuer le mapping et améliorer l'alignement de ces modèles, (ii) une utilisation croisée d'outils génériques et spécifiques à un domaine permettant de palier les problèmes de spécifications des vocabulaires liés à certains domaines, (iii) la conception d'une solution architecturelle décentralisée et communautaire pour permettre le passage à l'échelle des solutions de mappings manuels, qui mettent à profit des connaissances d'experts déjà présentes dans la recherche.

# Système de simulations de crises, basé sur une ontologie, pour la mise à l'abri des populations

**Jinfeng Zhong[1], Luyen Le Ngoc[2], Elsa Negre[1], Marie-Hélène Abel[2]**

*1. Paris-Dauphine University, PSL Research Universities, CNRS UMR 7243, LAMSADE, Paris, France.*
*jinfeng.zhong@dauphine.eu*

*2. Université de Technologie de Compiègne, CNRS, Heudiasyc, CS 60319 - 60203 Compiègne Cedex, France.*

*Cet article est une synthèse de : Zhong J, Ngoc LL, Negre E, Abel M-H. Ontology-based crisis simulation system for population sheltering management. SIMULATION. 2023;99(12).*

Les déséquilibres environnementaux conduisent à une augmentation des catastrophes naturelles. Si les impacts matériels et financiers sont importants, la sécurité des populations est primordiale. Les véhicules d'urgence (publics) sont utilisés pour évacuer les populations. Avec l'importance croissante de la participation citoyenne, de nombreuses personnes sont désormais désireuses de jouer un rôle actif lors d'une crise. Les actions qui en résultent doivent être coordonnées pour être optimisées et en accord avec les plans publics d'actions. Dans le contexte de l'évacuation des populations, les ressources publiques telles que les ambulances et les embarcations peuvent être limitées et ne pas être positionnées de manière optimale pour atteindre toutes les personnes dans le besoin. Dans de telles situations, il est nécessaire d'explorer des ressources d'évacuation alternatives : des citoyens peuvent être disposés à aider à l'évacuation en utilisant leurs propres véhicules.

Un défi important dans l'utilisation des ressources citoyennes pour l'évacuation des populations concerne leur diversité et leur description (contrairement aux ressources publiques référencées et décrites). Un autre défi consiste à identifier et à distribuer efficacement les conducteurs citoyens-volontaires[1]/véhicules disponibles dans les zones à risque.

La mobilisation et l'allocation de ressources conducteur/véhicule nécessitent calculs et optimisations en termes de nombre requis et temps de réponse estimé entre la localisation du véhicule et le point de secours. Nous voulons un système gérant

[1] Dans la suite de l'article, la ressource conducteur/véhicule fait exclusivement référence à un citoyen-volontaire capable de conduire son propre véhicule.

ces ressources conducteur/véhicule et proposant leur pertinente allocation pour l'évacuation des populations. Par conséquent, les deux problèmes clés sont : (P1) organiser les données/informations impliquées dans les ressources conducteur/véhicule et (P2) recommander des solutions sous les contraintes de capacité de ressources, de temps de réponse … Pour P1, une ontologie permettrait de normaliser la terminologie et modéliser les ressources conducteur/véhicule, les emplacements et les personnes affectées. Pour P2, identifier et distribuer efficacement les conducteurs/véhicules disponibles peut être abordé comme un problème de recommandation, dans lequel les ressources conducteur/véhicule sont des éléments recommandables aux décideurs publics pour évacuer les populations. Un système de recommandations à base de connaissances s'appuyant sur une ontologie peut aider à résoudre P2 en exploitant les exigences explicites des points de secours (nombre de personnes, niveau de priorité, ressources disponibles, …).

En conclusion, dans cet article, nous avons présenté un système de simulation de crise basé sur une ontologie pour la gestion de la mise à l'abri des populations. Après avoir construit une ontologie pour décrire et standardiser les ressources citoyennes, les situations de crise …, nous formulons le problème d'allocation des ressources comme un problème de recommandation, où les conducteurs/véhicules sont traités comme des éléments recommandables. Nous intégrons cela dans un système de simulation constitué d'un système d'aide à la décision à quatre niveaux : (1) « données » structure et stocke les données nécessaires, (2) « service » calcule le temps et la distance entre deux points géographiques, (3) « intelligence » calcule une liste de recommandations pour chaque point de secours et (4) « interactions » facilite les interactions entre les décideurs et le système de recommandation. Avec notre système, nous visons à aider les décideurs à se préparer à divers scénarios en optimisant l'allocation des ressources et en réduisant le temps nécessaire pour prendre des décisions.

Enfin, de nombreuses perspectives de recherche existent, comme (i) l'enrichissement de l'ontologie afin de traiter des situations plus complexes et en constante évolution, (ii) l'intégration de notre système dans un système complet de simulation comme (Laatabi et al., 2022) afin de réaliser plus d'expérimentations (e.g. avec GAMA (Grignard et al., 2013)), (iii) l'exploration des aspects sociotechniques (Land, 2000) notamment quant aux problèmes légaux d'utilisation des données d'assurance des citoyens-volontaires …

## Bibliographie

Grignard A, Taillandier P, Gaudou B et al. Gama 1.6: Advancing the art of complex agent-based modeling and simulation. *In PRIMA 2013: Principles and Practice of Multi- Agent Systems: 16th Int. Conference, Dunedin, New Zealand, December 1-6, 2013, pp. 117–131*

Laatabi A, Gaudou B, Hanachi C et al. Coupling agent-based simulation with optimization to enhance population sheltering. *In 19th Information Systems for Crisis Response and Management Conference (ISCRAM 2022)*.

Land F. *Evaluation in a socio-technical context.* In Organizational and social perspectives on information technology. Springer, 2000. pp. 115–126.

# Stratégies optimales pour l'analyse multidimensionnelle de contenus multilingues issus des réseaux sociaux

**Maxime Masson[1,2], Rodrigo Agerri[2], Christian Sallaberry[1], Marie-Noelle Bessagnet[1], Philippe Roose[1], Annig Le Parc Lacayrelle[1]**

*1. LIUPPA, E2S, Université de Pau et des Pays de l'Adour (UPPA)*
*2. Centre HiTZ – Ixa, Université du Pays Basque EHU/UPV*

RESUME. *L'influence grandissante des réseaux sociaux dans le domaine du tourisme souligne le besoin d'approches efficaces en traitement automatique du langage naturel (TALN) pour exploiter cette ressource. Toutefois, transformer des textes multilingues, informels et non structurés en connaissances structurées reste un défi, notamment à cause de la nécessité de données annotées pour l'entraînement des modèles. Cet article examine différentes techniques et modèles de TALN basés sur l'apprentissage pour optimiser les performances tout en réduisant le besoin de données annotées manuellement. Un nouveau jeu de données multilingues (français, anglais, espagnol) spécifique au tourisme a été créé, se concentrant sur la région du Pays Basque. Ce jeu de données inclut des tweets avec des annotations manuelles sur les entités nommées spatiales, les concepts thématiques touristiques et les sentiments. Une comparaison des méthodes de fine-tuning et d'apprentissage few-shot avec des modèles multilingues indique que les techniques few-shot peuvent produire de bons résultats avec peu d'exemples annotés. Les expérimentations menées sur ce jeu de données suggèrent la possibilité d'appliquer les méthodes de TALN à base d'apprentissage à divers domaines, tout en réduisant le besoin d'annotations manuelles et évitant les complexités des méthodes basées sur des règles.*

*Mots-clés : Tourisme, Apprentissage Few-Shot, Modèle de Langage Masqué (MLM), Multilinguisme, Science Sociale Informatique, Traitement Automatique du Langage Naturel*

## 1. Introduction

De nos jours, les réseaux sociaux se sont imposés comme des moyens de communication essentiels pour le partage d'opinions et d'expériences dans une variété de domaines, devenant ainsi une ressource précieuse pour les professionnels du tourisme tels que les offices de tourisme et les agences de voyage (Zeng *et al*., 2014). Toutefois, l'analyse de grandes quantités de données issues des réseaux sociaux représente un défi majeur (Maynard *et al*., 2012), en particulier pour extraire des connaissances structurées de textes non structurés. Ainsi, les acteurs du tourisme se tournent souvent vers les informaticiens et les linguistes pour l'extraction de connaissances, qui utilisent alors des techniques de Traitement Automatique du Langage Naturel (TALN). Le TALN est un outil puissant pour traiter et analyser les données textuelles, souvent employé pour automatiser des tâches telles que la

détection de sentiments, la reconnaissance d'entités nommées spatiales et l'extraction de concepts thématiques fins (Rosenthal *et al*., 2015). Les progrès récents dans ce domaine, notamment avec l'émergence de l'apprentissage profond et le développement des modèles de langage masqués (MLM), offrent des avantages significatifs par rapport aux méthodes traditionnelles basées sur des règles. Ces nouvelles techniques d'apprentissage automatique, plus adaptatives aux variations linguistiques (Min *et al.,* 2021), permettent une analyse plus dynamique et adaptative, en opposition aux approches basées sur des règles, souvent ad hoc et rigides. Toutefois, pour obtenir des résultats optimaux dans des domaines d'application spécifiques, les modèles d'apprentissage doivent préalablement subir une étape dite de « *fine-tuning* », c'est-à-dire un enrichissement avec des exemples annotés fortement liés au domaine concerné. Cela soulève deux questions récurrentes pour les chercheurs : (1) quelles sont les stratégies et modèles d'apprentissage les plus appropriés pour un domaine d'application et une tâche donnée, et (2) combien d'exemples annotés spécifiques au domaine sont nécessaires pour obtenir des résultats satisfaisants. L'annotation manuelle d'exemples est un processus souvent coûteux, fastidieux et chronophage, la majorité des chercheurs visent donc à obtenir les meilleurs résultats possibles tout en minimisant au maximum la quantité d'exemples annotés nécessaire.

Dans cet article, nous présentons une étude comparative sur les besoins en annotations pour obtenir de bonnes performances sur trois tâches d'extraction de connaissance communes appliquées au **domaine du tourisme**. Plus précisément, nous cherchons à savoir quelles stratégies d'apprentissage sont les meilleures pour minimiser le processus d'annotation manuelle des données et éviter les approches basées sur des règles, peu dynamiques. Nous supposons que parmi les modèles de langage masqués et les techniques d'entraînement existants, certains seront mieux adaptés à ce domaine précis. En effet, les messages des réseaux sociaux sont caractérisés par des textes informels courts, des erreurs grammaticales fréquentes et la présence d'emojis et hashtags. Nous nous concentrons sur les trois tâches d'extraction de connaissances suivantes : **la classification de la polarité des sentiments** (*classification de textes*), **la reconnaissance d'entités nommées spatiales** et **l'extraction de concepts thématiques fins** (*classification de tokens*).

Nos principales contributions sont les suivantes : (1) nous proposons un **nouveau jeu de données de tweets touristiques**. Ce jeu de données est multilingue (français, anglais et espagnol) et a été manuellement annoté au niveau du texte avec le *sentiment* (positif, négatif et neutre) et au niveau du token avec les *lieux* et les *concepts thématiques*. Ces derniers sont liés au *thésaurus du tourisme et des loisirs*[1] de l'Organisation Mondiale du Tourisme (OMT) ; (2) nous réalisons une **analyse comparative** entre des techniques de TALN basées sur des règles, le *fine-tuning* et l'apprentissage *few-shot* (Ma *et al.,* 2022) avec pour objectif d'établir quelle méthode est la plus efficace pour chaque tâche d'extraction de connaissances ; (3) finalement, nous expérimentons avec différentes méthodes d'échantillonnage de données pour déterminer **combien d'exemples annotés sont réellement**

[1] https://www.e-unwto.org/doi/book/10.18111/9789284404551

**nécessaires** pour obtenir des résultats compétitifs sur chacune des trois tâches. L'objectif est d'éviter aux chercheurs d'annoter manuellement une trop grande quantité de données par rapport à leurs besoins.

Cet article est structuré comme suit. Dans la section 2, nous passons en revue les techniques de TALN basées sur l'apprentissage profond les plus communément utilisés dans le domaine du tourisme. La section 3 couvre le processus de construction et d'annotation de notre jeu de données. La section 4 décrit la configuration expérimentale de notre étude. Dans la section 5, nous présentons une analyse comparative des différentes approches de TALN pour chacune des trois tâches décrites précédemment. Les résultats et les limitations sont discutés dans la section 6. Enfin, la section 7 présente les perspectives futures.

## 2. Travaux connexes

Dans le secteur en évolution constante du traitement automatique du langage naturel (TALN), l'une des avancées les plus significatives a été l'avènement des modèles de langage masqués (Masked Language Model ou MLM). Ces modèles, entraînés sur de vastes corpus textuels, capturent un large éventail de structures linguistiques, de nuances et de connaissances (Min et al., 2021). Ils offrent ainsi une amélioration significative des performances dans de nombreuses tâches de TALN, allant de la classification de textes ou de tokens aux systèmes de questions-réponses. Des modèles de langage masqués tels que BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018) et XLM-RoBERTa (Conneau et al., 2019) sont entraînés sur d'importantes quantités de données textuelles, englobant parfois des téraoctets d'informations issues de sources diverses telles que des livres, des articles et des sites Web. Grâce à ce processus d'entraînement étendu, ces modèles sont capables de comprendre les relations complexes entre mots, expressions et constructions textuelles. Ils capturent des signaux sémantiques, syntaxiques et contextuels, leur conférant la capacité de générer et de comprendre le langage naturel (Toporkov et Agerri, 2023).

### *2.1. Travaux antérieurs sur les réseaux sociaux dans le domaine du tourisme*

Une approche fréquemment utilisée consiste à ***fine-tuner*** (apprentissage par transfert) les modèles de langage masqués pour des tâches spécifiques au domaine (Sun *et al.,* 2019). Cela se fait en altérant les poids du modèle pour l'adapter à la nouvelle tâche, via un enrichissant de ce dernier avec des exemples annotés spécifiques au domaine étudié. Afin d'améliorer la précision, les modèles de langage masqués ont été affinés pour des tâches de classification de textes notamment la détection de *spam* dans les avis d'hôtels (Crawford et Khoshgoftaar, 2021) ou encore l'analyse des sentiments dans les avis touristiques (Enríquez *et al.,* 2022) ou dans le domaine du transport durable (Serna *et al.,* 2021). En REN (Reconnaissance d'Entités Nommées). Le *fine-tuning* de modèles de langage masqués a été employé pour extraire des informations de localisation de corpus touristiques (Cheng *et al.,* 2020). Enfin, les modèles de langage masqués ont démontré des résultats

prometteurs dans l'extraction de concepts thématiques, tels que l'identification de thèmes et sujets liés aux voyages dans des textes touristiques (Chantrapornchai et Tunsakul, 2021). L'un des principaux défauts de l'approche *fine-tuning* est qu'elle nécessite généralement un volume conséquent d'exemples annotés pour être efficace. Cela génère une charge de travail d'annotation importante pour les chercheurs.

### *2.2. Aborder le manque de données annotées spécifiques au domaine*

Face au défi des données annotées limitées, les techniques d'apprentissage ***few-shot*** ont gagné en popularité. Elles permettent aux modèles de langage masqués d'apprendre avec peu d'exemples annotés (de l'ordre de la dizaine), utiles lorsque les données abondantes font défaut.

Le Pattern-Exploiting Training (PET) est un exemple d'apprentissage *few-shot* dédié à la classification de texte, où le modèle est guidé par des descriptions de tâches en langage naturel et des phrases à trous (Schick et Schütze, 2020). Par exemple, pour classer les avis de films en fonction du sentiment prédominant qu'ils expriment, le modèle sera requêté (*prompt*) avec l'avis du film et la classification souhaitée : « *Le film était {?}* ». Le modèle essaierait alors de prédire le {?}, en choisissant parmi des options telles que « *génial* » ou « *décevant* ». Plusieurs travaux récents ont également appliqué ce principe à la classification de tokens, par exemple, EntLM (Ma *et al.,* 2022) pour la reconnaissance d'entités nommées. Dans les deux cas, les approches *few-shot* ont démontré leur efficacité pour obtenir des résultats satisfaisants avec peu d'exemples.

### *2.3. Ressources annotées existantes*

Bien que les données annotées disponibles publiquement pour le domaine du tourisme soient extrêmement réduites, il existe plusieurs corpus annotés dans d'autres domaines, qui pourraient être utilisés pour nos expérimentations.

Par exemple, le corpus **ESTER** (Galliano *et al.,* 2006) est une collection complète de transcriptions de radio françaises ; **AnCora** (Taulé *et al*., 2008) est un corpus annoté sur plusieurs niveaux (*principalement à partir de journaux)* pour le catalan et l'espagnol. Ces deux ressources sont annotées pour la reconnaissance d'entités nommées. En termes de ressources spécifiques aux réseaux sociaux, le **Broad Twitter Corpus** (Derczynski *et al.*, 2016) comprend des annotations sur les lieux, personnes et organisations, tandis que **Sentiment140** (Go *et al.,* 2009), **STS-Gold** (Saif *et al.,* 2013), et de nombreux autres jeux de données développés dans le cadre de tâches d'évaluation partagées comme **SemEval** (Rosenthal *et al.,* 2015), sont utilisés pour la classification de la polarité des sentiments. D'autres corpus incluent le jeu de données de dialogue **MultiWOZ** (Budzianowski *et al.*, 2018), le jeu de données **Stanford NLI** (Bowman *et al.,* 2015) pour l'inférence textuelle, et le corpus **Heldugazte** qui aide à catégoriser les tweets comme formels ou informels.

Ces jeux de données sont vastes mais très généraux et se concentrent souvent uniquement sur l'anglais, manquant ainsi d'informations contextuelles nécessaires

pertinentes au domaine du tourisme. Plus important encore, nous n'avons pas trouvé de jeu de données publiques annotés pour l'extraction de concepts thématiques fins dans le domaine du tourisme. Compte tenu de cela, nous avons décidé de construire notre propre jeu de données annoté.

## 3. Construction du jeu de données et processus d'annotation

Dans cette section, nous décrivons le processus de création d'un nouveau jeu de données multilingue composé de tweets liés au tourisme et annotés pour trois tâches d'extraction de connaissance pour des applications dans le domaine du tourisme : (1) la classification de la polarité des sentiments, (2) la reconnaissance des entités nommées spatiales, et (3) l'extraction de concepts thématiques fins (*basé sur le Thésaurus de l'OMT*).

Le jeu de données a été collecté via X en utilisant l'API Academic[2] et le processus de collecte a été mis en œuvre en utilisant une méthodologie que nous avons conçue pour la construction de jeux de données. Cette méthodologie est à la fois générique, itérative et incrémentale (voir Masson *et al*., 2022 pour plus de détails). Plusieurs itérations ont été réalisées, chacune avec un filtrage successif correspondant aux *dimensions* cibles du jeu de données : spatiale (zone de la côte Basque française, coordonnées et toponymes), temporelle (été 2019, horodatage) et thématique (domaine du tourisme tel que définis par le thésaurus de l'Organisation Mondiale du Tourisme). Pour éviter un bruit excessif, chaque itération a été suivie d'un *feedback* humain pour ajuster et équilibrer les critères de filtrage. De même, nous avons exclu les utilisateurs professionnels et institutionnels car nous sommes principalement intéressés par la compréhension du comportement des touristes en tant qu'individu, et non par l'analyse de contenus promotionnels ou institutionnels.

Le jeu de données final comprend 27 379 tweets, parmi lesquels **2 961 tweets** provenant de 624 utilisateurs ont été sélectionnés pour l'annotation et l'utilisation dans nos expérimentations. Ces tweets spécifiques (2 961) ont été sélectionnés car ils ont été examinés manuellement pour s'assurer qu'ils concernent le tourisme et **sont émis par des touristes** (e.g., nous avons déterminé que seuls 624 utilisateurs du jeu de données initial sont des touristes, représentant environ 10 % des 27 379 tweets initiaux), plutôt que par des professionnels du tourisme ou des médias traitant du tourisme, entre autres. Ce choix privilégie la qualité des tweets par rapport à la quantité. Le jeu de données est multilingue et inclut une variété de tweets en anglais, français et espagnol. Le déséquilibre entre les différentes langues reflète la réalité de l'utilisation des réseaux sociaux sur la côte basque française.

Le Tableau 1 montre la répartition linguistique du jeu de données et la subdivision des tweets pour l'apprentissage (60% pour l'entraînement, 20% pour le développement, 20% pour le test). Cette division maintient un équilibre entre le nombre d'utilisateurs et de langues dans chaque jeu de données.

[2] https://developer.twitter.com/en/use-cases/do-research/academic-research (fin du service en avril 2023)

*Tableau 1. Répartition du jeu de données collecté par langue – Tweets (Utilisateurs)*

| | Tous | Français | Anglais | Espagnol |
|---|---|---|---|---|
| Train | 1 652 (503) | 1 297 (391) | 283 (129) | 82 (32) |
| Dev | 619 (300) | 450 (213) | 99 (66) | 70 (31) |
| Test | 680 (431) | 401 (273) | 102 (100) | 177 (93) |

### *3.1 Annotation du sentiment*

Le processus d'annotation des 2 961 tweets a été effectué **semi-automatiquement**, comme illustré dans la Figure 1. Pour faciliter le travail des annotateurs humains, 1 299 tweets des divisions *dev* et *test* ont été **pré-annotés automatiquement** (au niveau du texte) en utilisant les 5 modèles de langage dédiés à la prédiction du sentiment décrits dans le Tableau 2.

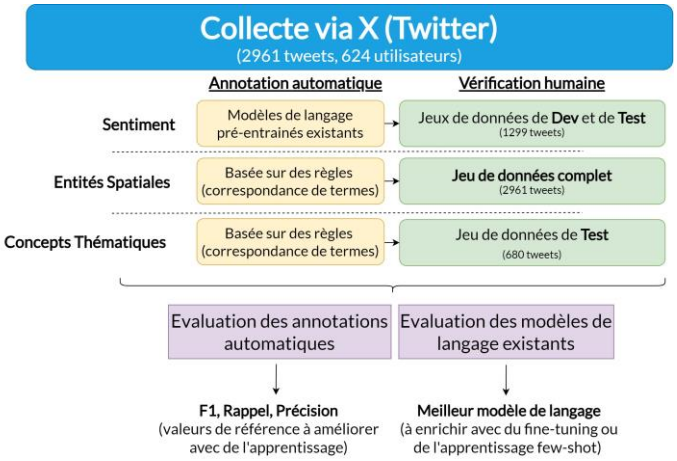



*Figure 1. Processus de construction et d'annotation du jeu de données*

Ces annotations ont ensuite été révisées manuellement par les annotateurs humains. Chaque tweet a été évalué par deux annotateurs pour mesurer la concordance (via le coefficient *kappa de Cohen*), assurant ainsi la qualité des annotations. Nous avons obtenu $\kappa = 0.79$ pour les tweets en français, $\kappa = 0.75$ pour l'espagnol, et $\kappa = 0.67$ pour l'anglais, indiquant un accord fort. Les divergences ont été résolues par discussion collaborative.

L'étape suivante a consisté à évaluer la performance des 5 modèles de langage choisis pour l'annotation automatique, en les comparant aux annotations révisées par les annotateurs humains. Le Tableau 2 montre que le modèle XLM-T Sentiment (Barbieri *et al.,* 2022) qui est déjà *fine-tuné* avec des données multilingues

sentimentales provenant de domaines variés, a obtenu les meilleurs résultats en moyenne pour les trois langues. Ce modèle (XLM-T Sentiment) a été utilisé pour annoter automatiquement le jeu de données *train*, que nous avons révisé manuellement pour nos expérimentations.

*Tableau 2. Précision de modèles de prédiction du sentiment existants (jeu de test)*

| **Langue des tweet** | **Barbieri *et al.*, 2020** | **Perez *et al.*, 2021** | **Seethal *et al.*, 2023** | **Hartmann *et al.*, 2023** | **Barbieri *et al.*, 2022** |
|---|---|---|---|---|---|
| Français | 0.56 | 0.45 | 0.43 | 0.47 | **0.82** |
| Espagnol | 0.71 | 0.64 | 0.61 | 0.34 | **0.83** |
| Anglais | **0.81** | **0.81** | 0.71 | 0.66 | **0.80** |

### *3.2 Annotation des lieux et des concepts thématiques touristiques*

Nos tâches suivantes s'intéressent à la reconnaissance d'entités nommées spatiales et à l'extraction de concepts thématiques détaillés dans le domaine du tourisme. Contrairement à la détection de sentiments, qui relève de la classification de textes, ces tâches s'inscrivent dans le cadre de la classification de tokens. Autrement dit, chaque token dans les textes est annoté individuellement.

Avant d'expérimenter des méthodes d'apprentissage automatique, nous avions développé une méthode d'annotation basée sur la correspondance de termes. La précision et le taux de rappel obtenus par cette méthode serviront de référence à améliorer avec des méthodes par apprentissage. Pour détecter les localisations, nous avons utilisé 625 toponymes locaux issus d'Open Street Map, incluant des villes, des points d'intérêt et des repères. Quant aux concepts thématiques, ils ont été identifiés en utilisant leurs étiquettes et synonymes dans le thésaurus de l'Organisation Mondiale du Tourisme, qui recense 1 494 concepts liés au tourisme. Un prétraitement des tweets (lemmatisation, mise en minuscules, suppression des URL, décomposition des hashtags) a été effectué pour faciliter la correspondance des termes. Nous avons appliqué cette méthode pour annoter l'ensemble du jeu de données (*train*, *dev* et *test*). Les annotations générées automatiquement ont par la suite été révisées manuellement par des annotateurs humains. Cette révision a été appliquée sur l'ensemble du jeu de données pour les entités spatiales, tandis que pour les concepts thématiques, elle a été limitée au jeu de *test*. Pour les concepts thématiques, la méthode par correspondance de termes a détecté 315 classes de concept unique (sur les 1 494 concepts inclus dans le thésaurus de l'OMT), donnant ainsi une tâche de classification de tokens de granularité très fine. Nous n'avons effectué les révisions manuelles que sur le jeu de test, car annoter 315 classes de concepts est une tâche complexe et demandant un temps conséquent.

L'accord entre annotateurs a été mesuré sur un échantillon aléatoire de 100 tweets. Concernant les entités spatiales, le coefficient Kappa atteint 0,91 pour les correspondances exactes, c'est-à-dire lorsque tous les tokens constituant une entité correspondent (par exemple, dans le cas de la ville de *New York*, les tokens *New* et

*York*). Pour les correspondances partielles, où une entité est identifiée mais présente des tokens manquants ou additionnels (*New* seulement sans le *York*), le coefficient est de 0,93. Ces valeurs témoignent d'un accord quasi parfait entre les annotateurs.

*Tableau 3. Performance de la méthode d'annotation par correspondance de termes*

| **Reconnaissance d'entités nommées spatiales** | **Rappel** | **Précision** | **Mesure F1** |
|---|---|---|---|
| Correspondance Exacte | **0.692** | 0.722 | 0.707 |
| Correspondance Partielle | 0.780 | 0.814 | 0.796 |
| **Extraction de concepts thématiques fins** | **Rappel** | **Précision** | **Mesure F1** |
| Correspondance Exacte | **0.746** | **0.952** | 0.836 |
| Correspondance Partielle | 0.747 | 0.953 | 0.837 |

Les performances de la méthode par correspondance de termes appliquée au jeu de test sont rapportées dans le Tableau 3. Ces résultats serviront de référence pour comparer avec les différentes approches par apprentissage dans la section expérimentale. Nous constatons que, pour les localisations, les performances ne sont pas satisfaisantes, notamment en termes de rappel (**0.692**). Cependant, la méthode par correspondance de termes se distingue nettement par sa précision sur l'extraction de concepts thématiques fins (**0.952**). Bien que cela constitue une valeur de référence solide, le rappel reste relativement faible (**0.746**), ce qui signifie que de nombreux concepts thématiques ne sont pas détectés. Ainsi, notre principal objectif est désormais de déterminer si les techniques d'apprentissage automatique peuvent égaler ou surpasser cette méthode tout en minimisant la quantité d'annotations manuelles, en particulier pour l'extraction de concepts thématiques fins.

## 4. Protocole expérimental

L'expérimentation se concentre sur les trois tâches présentées précédemment : classification de la polarité des sentiments (section 4.1), reconnaissance d'entités nommées spatiales et extraction de concepts thématiques fins (section 4.2) afin de déterminer quelles approches sont les plus efficaces et avec quelle quantité de données d'entraînement. La Figure 2 donne un aperçu de notre configuration expérimentale avec (1) les modèles de langage, (2) les méthodes d'échantillonnage et (3) les méthodes d'apprentissage automatique utilisées pour chaque tâche. Les données d'entrainement sont échantillonnées en utilisant deux méthodes :

- **Échantillonnage k-shot :** sélection d'un nombre précis d'exemples pour chaque classe d'annotation. Par exemple, dans le cas de la classification de la polarité des sentiments, si l'on souhaite réaliser un échantillonnage 5-shot, il faudra 15 exemples (5 positifs, 5 négatifs et 5 neutres). Pour nos expérimentations, nous avons utilisé les valeurs de k suivantes : 5, 10, 20, 30, 40, 50, et 100 exemples par classe.
- **Échantillonnage par pourcentage** : utilisation d'un pourcentage précis du jeu de données. Nous avons successivement utilisé 5%, 10%, 20%, 30%, 40%,

50%, 60%, 70%, 90%, et 100% du jeu de données d'entrainement, tout en essayant de maintenir la distribution originale des classes d'annotation, y compris les étiquettes O (l'étiquette O est attribuée aux tokens qui ne sont ni des lieux ni des entités thématiques).

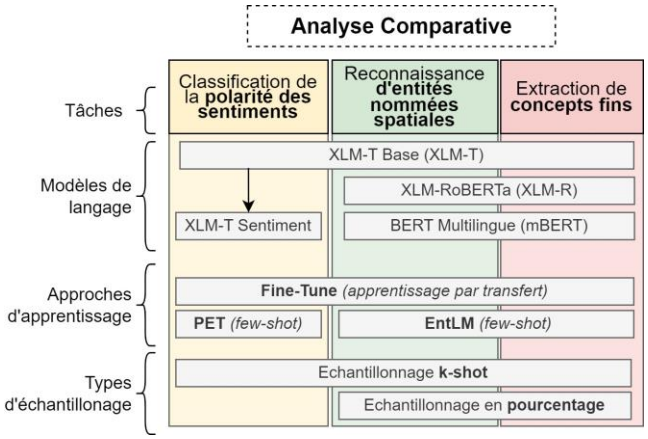



*Figure 2. Configuration expérimentale de l'étude*

### *4.1 Classification de textes - Classification de la polarité des sentiments*

Sur la base des résultats rapportés par le Tableau 2, nous avons choisi le modèle de langage XLM-T (Barbieri *et al.*, 2022) pour nos expérimentations sur la classification de la polarité des sentiments. XLM-T est basé sur XLM-RoBERTa (Conneau *et al.,* 2019), mais entraîné sur un corpus de 198 millions de tweets comprenant 15 langues. Cette version du modèle est spécialement conçue pour gérer les caractéristiques uniques des tweets et plus généralement des publications sur les réseaux sociaux (longueur limitée, langage informel, présence d'emojis, etc.). Nous utilisons 2 variantes :

- La **version de base,** dénommée XLM-T (Barbieri *et al.,* 2022), qui est un modèle de langage masqué (MLM). Elle permet de prédire le prochain mot.
- **XLM-T fine-tuné pour la classification de la polarité des sentiments** (dénoté XLM-T Sentiment). Cette variante permettant la prédiction du sentiment a été préalablement *fine-tuné* en utilisant 24 264 tweets couvrant 8 langues différentes (incluant le français, l'anglais, et l'espagnol). Cependant, ces tweets couvrent un large éventail de domaines n'incluant pas le tourisme.

Ces deux modèles vont ensuite être *fine-tunés* ou utilisés dans des approches de requêtage *few-shot* avec plusieurs échantillons de notre jeu de données d'entrainement générés via les deux méthodes d'échantillonnage décrites précédemment. Nous utilisons les méthodes d'apprentissage suivantes.

- **Fine-Tuning** : les hyperparamètres ont été déterminés en testant toutes les combinaisons possibles (recherche en grille).
- **Pattern-Exploiting Training (PET)** (voir Schick et Schütze, 2020) : il s'agit d'une approche d'apprentissage *few-shot* pour la classification de textes. Elle est basée sur le concept de phrases à trous (*cloze*). Dans notre cas d'utilisation, la requête (*prompt*) envoyée au modèle est formulée comme suit : « *Le sentiment dominant exprimé dans le texte suivant : [Tweet] est {?}* ». Le modèle de langage masqué utilisé tentera alors de remplir le {?} avec le sentiment approprié à partir d'une liste d'étiquettes possibles : positif, négatif ou neutre.

En comparant ces deux méthodes d'apprentissage et en évaluant leur efficacité avec différentes quantités d'exemples annotés, nous visons à mieux comprendre les exigences minimales en termes de données pour obtenir des résultats fiables pour de la classification de la polarité des sentiments dans le domaine du tourisme.

### *4.2 Classification de tokens – Reconnaissance d'entités nommées spatiales et extraction de concepts thématiques fins*

Pour la classification de tokens, englobant à la fois les lieux et les thèmes, nous adoptons une démarche similaire. Toutefois, dans le cadre de l'apprentissage *few-shot*, nous optons pour la méthode EntLM (Ma *et al.,* 2022), spécifiquement conçue pour la catégorisation de tokens. En complément du modèle XLM-T, nous intégrons également deux autres modèles à notre étude : XLM-RoBERTa (XLM-R) et BERT multilingue (mBERT, Devlin et al., 2018), ce dernier étant le modèle par défaut utilisé par EntLM. Pour rappel, l'un des objectifs de cette étude comparative est de déterminer la quantité minimale d'exemples annotés nécessaires pour justifier le passage de méthodes basées sur des règles (comme celle par correspondance de termes) rigides à des méthodes d'apprentissage plus dynamique mais nécessitant des exemples annotés pour l'entrainement. Plus spécifiquement, nous cherchons à déterminer le point de basculement à partir duquel les avantages de l'utilisation des techniques par apprentissage l'emportent sur leurs exigences en matière d'exemples.

Pour la classification de la polarité des sentiments, nous rapportons des résultats de précision, tandis que pour la classification de tokens (lieux et thèmes), nous utilisons la métrique micro-F1 calculée au niveau du segment tel que défini dans la tâche partagée CoNLL 2002 (Tjong Kim Sang et Erik, 2002). Tous les résultats rapportés sont la moyenne de trois exécutions initialisées aléatoirement.

## 5. Résultats

**Classification de la polarité des sentiments**. Les résultats obtenus pour la tâche de classification de la polarité des sentiments sont rapportés dans la Figure 3. Pour rappel, cette tâche consiste à annoter chaque tweet avec une des classes d'annotation suivantes : *positif*, *négatif* ou *neutre*. L'axe *x* représente le nombre d'exemples annotés utilisés pour l'entrainement, tandis que l'axe *y* indique les scores

de précision obtenus en utilisant deux méthodes d'apprentissage différentes : le fine-tuning (F-T) et le *Pattern-Exploiting Training* (PET, une méthode de type *few-shot*). Deux modèles de langage sont utilisés XLM-T et XLM-T Sentiment. Les résultats montrent que le *fine-tuning* du modèle XLM-T Sentiment est plus performant que l'apprentissage *few-shot* (PET). Cette observation suggère qu'un *fine-tuning* réalisé sur un vaste ensemble de données multilingues pour la détection de sentiments, même issues de domaines très différents, contribue significativement à l'amélioration des performances sur des données liées au tourisme (Figure 3, (b)). Cela se traduit par une nette amélioration des résultats dans des situations où les données sont peu abondantes. Le modèle XLM-T Sentiment, après *fine-tuning*, atteint une efficacité optimale avec seulement 10 exemples et parvient à égaler les performances obtenues en utilisant l'ensemble des exemples avec un entraînement basé sur 5 exemples seulement.

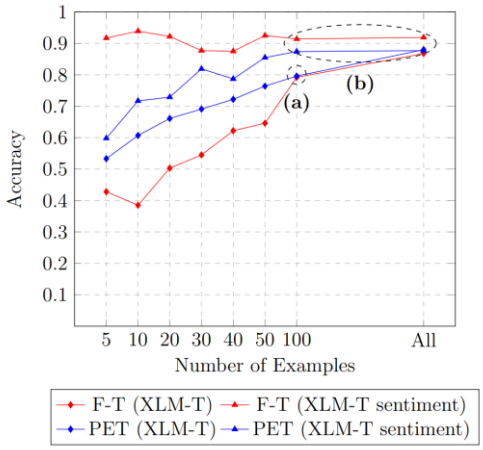



*Figure 3. Classification de la polarité des sentiments - Echantillonnage k-shot*

Lorsqu'on se concentre sur les approches qui exploitent uniquement nos propres données d'entraînement, l'apprentissage *few-shot* (PET) surclasse le fine-tuning du modèle XLM-T (jusqu'à 100 exemples annotés, voir Figure 3 : (a)). Cette observation met en évidence l'efficacité de PET, capable d'obtenir une précision importante même avec un nombre très restreint d'exemples (en revanche, la performance de PET n'atteint pas celle obtenue par le fine-tuning du modèle XLM-T Sentiment, qui tire avantage d'un pré-entraînement sur un vaste corpus externe composé de milliers de tweets). En résumé, nous pouvons tirer deux enseignements significatifs de ces résultats: (1) lors de l'utilisation d'un modèle de sentiment déjà pré-entrainé comme XLM-T Sentiment, un jeu de données d'entrainement contenant aussi peu que 10 exemples est suffisant pour obtenir de bonnes performances pour faire de la classification de la polarité des sentiments dans le domaine du tourisme, ajouter plus d'exemples ne semble pas améliorer significativement la précision et (2)

lors de l'emploi d'un modèle de langage masqué (MLM) comme XLM-T, PET semble être un choix préférable pour des scénarios à faible disponibilité de données, étant donné qu'une performance quasi optimale peut être atteinte avec 50 exemples.

**Reconnaissance d'entités nommées spatiales**. La Figure 4 montre la performance de la reconnaissance d'entités nommées spatiales avec les deux méthodes d'échantillonnage (k-shot et pourcentage). Les versions *fine-tunées* de trois modèles : XLM-T, XLM-RoBERTa (XLM-R), et mBERT ont été comparées à EntLM, une méthode d'apprentissage *few-shot* pour les tâches de classification de tokens basée sur un modèle BERT multilingue. En comparant les résultats en utilisant les deux méthodes d'échantillonnage différentes (cf. Figure 4), nous pouvons observer que lors de l'utilisation de toutes les données d'entraînement, les quatre méthodes obtiennent des résultats relativement comparables. Cependant, dans un contexte de faible disponibilité des données, la méthode EntLM nécessite moins d'exemples annotés. En d'autres termes, le BERT multilingue *fine-tuné* ne commence à surpasser EntLM qu'à partir de l'utilisation de 30% des données d'entraînement. Globalement, le *fine-tuning* n'est compétitif qu'en utilisant l'échantillonnage par pourcentage. Nous pensons que cela pourrait être dû à la faible quantité de tokens n'étant pas des localisations générées par l'échantillonnage k-shot. En revanche, EntLM se comporte de manière assez robuste en utilisant un plus petit nombre d'exemples avec les deux méthodes d'échantillonnage.

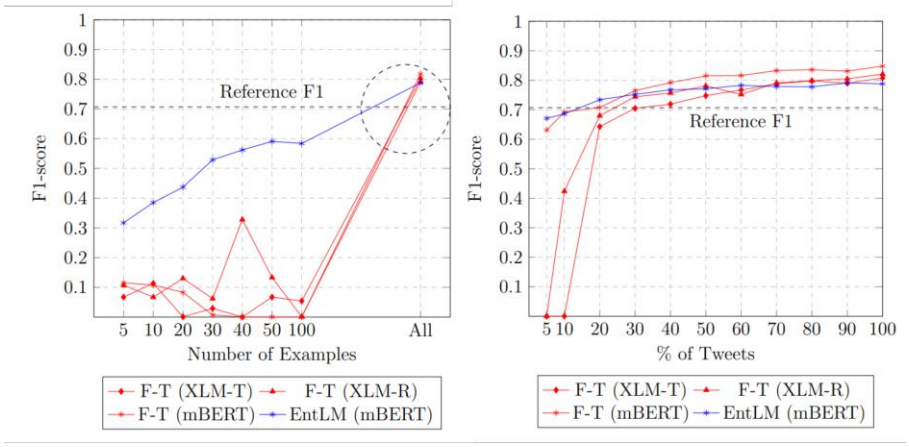



*Figure 4. Reconnaissance d'entités nommées spatiales – Echantillonnage des données d'entrainement par k-shot (gauche) et pourcentage (droite)*

Nous constatons également que le BERT multilingue *fine-tuné* et EntLM sont très nettement meilleurs comparés à l'approche basée sur des règles lorsqu'ils utilisent seulement environ 13% des tweets pour l'entraînement (~ 200 tweets). De manière surprenante, les modèles généraux (mBERT et XLM-R) obtiennent de meilleures performances que ceux entraînés sur des données X/Twitter (XLM-T).

Ce résultat montre que dans certains cas, les modèles généraux tendent à mieux s'appliquer à des domaines spécifiques que des modèles parfois trop spécialisés.

**Extraction de concepts thématiques fins**. C'est dans l'extraction de concepts thématiques fins, présentée dans la Figure 5, que l'approche d'apprentissage *few-shot* EntLM se distingue le plus. Pour cette tâche, impliquant la catégorisation de tokens dans un inventaire de 315 classes, EntLM se montre très compétitif. Ainsi, avec juste 5 exemples par classe (paramétrage 5-shot), il obtient un score F1 de 0.760. De même, il égale les résultats de l'approche par correspondance de termes avec un entraînement sur seulement 50 exemples. Ces scores indiquent une forte capacité à identifier avec précision les concepts touristiques, comme en témoignent les valeurs de précision élevées allant de 0.80 à 0.913. Bien que les résultats obtenus avec l'approche par correspondance de termes soient similaires, EntLM est légèrement supérieur en termes de rappel tout en étant légèrement moins bon en précision. Néanmoins, la performance d'EntLM est prometteuse pour éviter l'effort d'annotation manuelle ou le développement complexe d'approches basées sur des règles pour des tâches de classification de tokens fines spécifiques à un domaine.

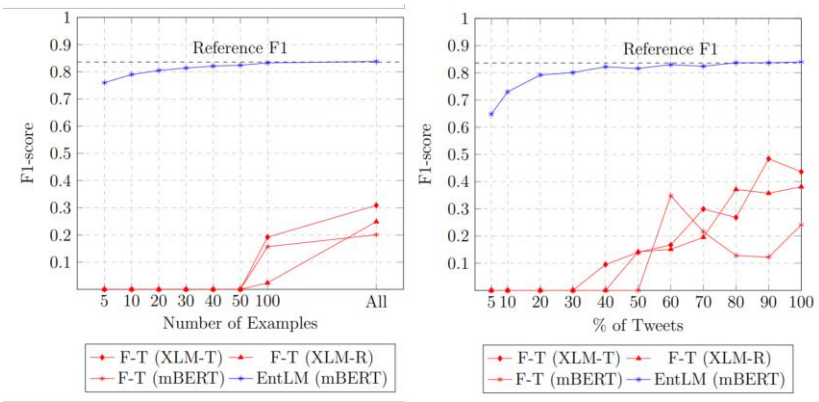



*Figure 5. Extraction de concepts thématiques fins – Echantillonnage des données d'entrainement par k-shot (gauche) et pourcentage (droite)*

## 6. Discussion et limitations

Après ces expérimentations, penchons-nous sur les conclusions obtenues pour discuter des principaux éléments ayant émergé. Nous aborderons également les limitations potentielles qui pourraient avoir affecté les résultats. Nos résultats montrent que le fine-tuning sur des modèles déjà pré-entraînés, comme XLM-T Sentiment, peut s'avérer très efficace pour la détection de sentiments, même avec un faible volume de données spécifiques au domaine. Cette observation souligne l'importance d'un pré-entraînement riche et varié pour améliorer la performance des modèles dans des contextes de données limitées. Cependant, il est crucial de souligner que l'efficacité de cette approche dépend fortement de la disponibilité de

données pré-entraînées pertinentes et de la capacité du modèle à s'adapter au contexte du tourisme.

Pour la tâche de reconnaissance d'entités nommées spatiales (une classe d'annotation *localisation*, mais beaucoup de mots labels associé à cette dernière), nos expérimentations indiquent que les méthodes d'apprentissage *few-shot*, comme EntLM, peuvent rivaliser avec des techniques de fine-tuning plus gourmandes en données. Cette observation suggère que des approches d'apprentissage plus légères (comme celles basées sur le principe du *few-shot* comme EntLM) peuvent être suffisantes pour traiter des tâches de NER spécifiques, en particulier dans des contextes où les données annotées sont rares ou coûteuses à obtenir.

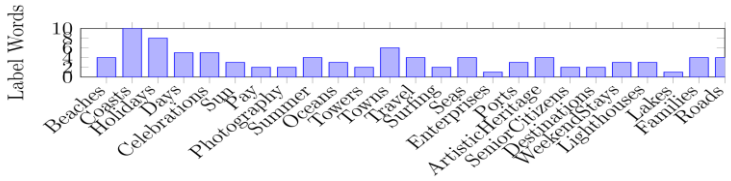



*Figure 6. Nombre de mots labels pour les concepts les plus fréquents du corpus*

L'extraction de concepts thématiques fins (315 classes correspondant aux concepts du thésaurus de l'OMT, chacun ayant un nombre restreint de mots labels, voir Figure 6) en revanche, présente des défis uniques en raison de la granularité et de la spécificité des classes impliquées. Bien que la méthode *few-shot* EntLM ait démontré une efficacité certaine, il est important de reconnaître que la précision de cette approche dépend fortement de la représentativité des exemples d'entraînement et de leur alignement avec les concepts thématiques du domaine du tourisme.

Les limitations de cette étude incluent la taille relativement restreinte de notre corpus spécifique au tourisme, qui pourrait influencer la généralisabilité de nos résultats. Bien que notre corpus soit multilingue et annoté avec soin, l'étendre à d'autres langues ou contextes touristiques pourrait fournir des indications supplémentaires sur l'adaptabilité des modèles d'apprentissage dans des contextes variés. De plus, bien que nous ayons concentré nos efforts sur des tâches spécifiques liées au tourisme, les méthodes et conclusions pourraient nécessiter une validation supplémentaire dans d'autres domaines d'application. Notre étude contribue à une meilleure compréhension des stratégies optimales pour l'analyse de données multilingues issues des réseaux sociaux dans le domaine du tourisme. Nos résultats soulignent l'importance de choisir la bonne méthode d'apprentissage en fonction de la spécificité de la tâche, de la disponibilité des données et de la nécessité d'annotations manuelles. Des recherches futures pourraient explorer l'extension de ces techniques à d'autres domaines ou l'intégration de sources de données diversifiées pour enrichir la capacité d'analyse des modèles d'apprentissage profond.

## 7. Conclusion et perspectives

Cet article propose une étude comparative de plusieurs techniques d'apprentissage sur 3 tâches d'extraction de connaissance : la classification de la polarité des sentiments, la reconnaissance d'entités nommées spatiales et l'extraction de concepts thématiques fins dans le domaine du tourisme sur les réseaux sociaux. L'objectif est de déterminer la meilleure stratégie pour obtenir des résultats performants tout en réduisant au maximum les annotations manuelles et le développement de méthodes basées sur des règles, souvent complexe. Pour cela, un nouveau jeu de données multilingue spécifique au tourisme sur *X/Twitter* a été créé. Ce jeu de données, qui sera rendu public dans les prochains mois, comprend des annotations au niveau du texte sur le sentiment et des tokens sur les lieux et concepts thématiques. Les résultats de notre étude confirment que l'apprentissage *few-shot* est particulièrement efficace pour ces tâches avec peu d'exemples annotés. Ce résultat est pertinent non seulement pour le développement d'applications spécifiques au tourisme mais aussi pour d'autres domaines nécessitant du TALN. Des recherches supplémentaires sont cependant nécessaires pour valider la généralisabilité de nos résultats dans d'autres domaines d'application. La prochaine étape du projet est de présenter ces résultats aux acteurs de l'industrie du tourisme, notamment à travers des tableaux de bord dynamiques mettant en évidence les entités extraites des réseaux sociaux, en lien avec des données contextuelles comme le sentiment.

## Bibliographie

Barbieri F., Espinosa Anke L., Camacho-Collados J. (2022). XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond. *LREC 2022*, p. 258-266.

Barbieri, F., Camacho-Collados, J., Neves, L., Espinosa-Anke, L. (2020). Tweeteval: Unified benchmark and comparative evaluation for tweet classification. arXiv:2010.12421

Bowman S.R., Angeli G., Potts C., Manning C.D. (2015). A large annotated corpus for learning natural language inference. *arXiv preprint* arXiv:1508.05326.

Budzianowski P., Wen T.-H., Tseng B.-H., Casanueva I., Ultes S., Ramadan O., Gašić M. (2018). MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling. *EMNLP 2018*, novembre 2018, p. 5016-5026.

Chantrapornchai C., Tunsakul A. (2021). Information extraction on tourism domain using SpaCy and BERT. *ECTI Transactions on Computer and IT*, vol. 15, n° 1, p. 108-122.

Cheng X., Wang W., Bao F., Gao G. (2020). MTNER: A Corpus for Mongolian Tourism Named Entity Recognition. *CCMT 2020*, October 10-12, 2020, Springer, p. 11-23.

Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., Grave E., Ott M., Zettlemoyer L., Stoyanov V. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint* arXiv:1911.02116.

Crawford M., Khoshgoftaar T.M. (2021). Using inductive transfer learning to improve hotel review spam detection. IRI 2022*,* IEEE, p. 248-254.

Devlin J., Chang M.-W., Lee K., Toutanova K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint* arXiv:1810.04805.

Derczynski L., Bontcheva K., Roberts I. (2016). Broad Twitter Corpus: A Diverse Named Entity Recognition Resource. *COLING 2016: Technical Papers*, p. 1169-1179.

Enríquez M.P., Mencía J.A., Segura-Bedmar I. (2022). Transformers Approach for Sentiment Analysis: Classification of Mexican Tourists Reviews from TripAdvisor.

Galliano S., Geoffrois E., Gravier G., Bonastre J.-F., Mostefa D., Choukri K. (2006). Corpus description of the ESTER Evaluation Campaign for the Rich Transcription of French Broadcast News. *LREC*, Citeseer, p. 139-142.

Go A., Bhayani R., Huang L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report*, Stanford, vol. 1, n° 12, p. 2009.

Hartmann, J., Heitmann, M., Siebert, C., Schamp, C. (2023). More than a feeling: Accuracy and application of sentiment analysis. *Int. Journal of Research in Marketing*, 40(1):75–87

Pérez, J. M., Giudici, J. C., Luque, F. (2021). pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks. *arXiv e-prints*, arXiv-2106.

Ma R., Zhou X., Gui T., Tan Y., Li L., Zhang Q., Huang X. (2022). Template-free Prompt Tuning for Few-shot NER. *NAACL 2022: Human Language Technologies*, p. 5721-5732.

Masson M., Sallaberry C., Agerri R., Bessagnet M.-N., Roose P., Le Parc Lacayrelle A. (2022). A Domain-Independent Method for Thematic Dataset Building from Social Media: The Case of Tourism on Twitter. *Web Information Systems Engineering*, p. 11-20.

Maynard D., Bontcheva K., Rout D. (2012). Challenges in developing opinion mining tools for social media, *Workshop Programme*, p. 15.

Min B., Ross H., Sulem E., Veyseh A.P.B., Nguyen T.H., Sainz O., Agirre E., Heinz I., Roth D. (2021). Recent advances in natural language processing via large pre-trained language models: A survey. *arXiv preprint* arXiv:2111.01243.

Rosenthal S., Nakov P., Kiritchenko S., Mohammad S., Ritter A., Stoyanov V. (2015). SemEval-2015 Task 10: Sentiment Analysis in Twitter. *SemEval 2015*, p. 451-463.

Saif H., Fernandez M., He Y., Alani H. (2013). Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold.

Schick T., Schütze H. (2020). Exploiting cloze questions for few shot text classification and natural language inference. *arXiv preprint* arXiv:2001.07676.

Serna A., Soroa A., Agerri R. (2021). Applying Deep Learning Techniques for Sentiment Analysis to Assess Sustainable Transport. *Sustainability*, vol. 13, n° 4, article 2397.

Seethal (2023). Sentiment analysis generic dataset.. *En ligne le 23 mars 2023.*

Sun, C., Qiu, X., Xu, Y., Huang, X. (2019). How to fine-tune bert for text classification? *Chinese Computational Linguistics: 18th China National Conference, CCL 2019*.

Taulé, M., Martí, M. A., Recasens, M. (2008). Ancora: Multilevel annotated corpora for Catalan and Spanish. *LREC*. Vol. 2008, pp. 96-101.

Tjong Kim Sang, Erik F. (2002). Introduction to the CoNLL-2002 Shared Task. *CoNLL 2002*

Toporkov O., Agerri R. (2023). On the Role of Morphological Information for Contextual Lemmatization. *arXiv preprint,* vol. 2302.00407.

Zeng B., Gerritsen R. (2014). What do we know about social media in tourism? A review. Tourism Management Perspectives, vol. 10, Elsevier, p. 27-36.

# Un modèle de réification pour les graphes de propriétés : application à l'intégration de données et de connaissances multi-sources relatives aux handicaps

**Selsebil Benelhaj-Sghaier, Annabelle Gillet, Éric Leclercq**

*Laboratoire d'Informatique de Bourgogne - EA 7534*
*Université de Bourgogne Franche-Comté*
*Dijon, France*
*Selsebil_Ben-El-Haj-Sghaier@etu.u-bourgogne.fr,{prenom}.{nom}@u-bourgogne.fr*

RÉSUMÉ. *La diversité des données et des applications des systèmes d'information fait apparaître le besoin d'associer de la connaissance aux données afin de pouvoir leur ajouter une signification lors de leur manipulation. Les graphes de connaissances sont une solution flexible à ce besoin. Le modèle de graphe utilisé par les graphes de connaissances définit leur expressivité, c'est-à-dire les constructions qu'il est possible d'utiliser pour représenter la connaissance. Cependant, les modèles de graphe actuels possèdent des limites en terme d'expressivité, et ne permettent pas de représenter certaines relations complexes. Nous proposons d'étendre le modèle de graphe de propriétés afin d'améliorer son expressivité, en ajoutant la possibilité d'abstraire un sous-graphe sous la forme d'un nœud qui peut avoir des étiquettes, des propriétés et des liens vers d'autres nœuds, abstraits ou non. Cela permet de définir différents niveaux d'abstraction dans les graphes de connaissances et de leur associer des métadonnées.*

ABSTRACT. *The diversity of data and applications of information systems reveals the need to associate knowledge and data to add a signification when manipulating data. Knowledge graphs are a flexible solution to this need. The graph model used by a knowledge graph defines its expressivity, namely the constructions that can be used to represent knowledge. However, modern graph models have a limited expressivity, and cannot represent complex relationships. We propose to extend the property graph model to improve its expressivity, by adding the capability to abstract a subgraph as a node, that can have labels, properties and links toward other nodes, whether they are abstract or not. This mechanism allows to define multiple levels of abstraction with metadata in a knowledge graph.*

MOTS-CLÉS : *Graphe de connaissances, Graphe de propriétés, Abstraction multi-niveaux, Relations complexes.*

KEYWORDS: *Knowledge graph, Property graph, Multi-level abstraction, Complex relationships.*

## 1. Introduction

Les systèmes d'information doivent faire face à des données toujours de plus en plus volumineuses et variées, tout en mettant à disposition des utilisateurs une multitude d'applications se basant sur ces données. Chaque application, en fonction de ses objectifs, a besoin d'avoir accès à la signification des données qu'elle manipule, qui est alors apportée par des connaissances externes associées aux données.

Dans un contexte de données mutli-sources hétérogènes ayant recours à des connaissances variées pour supporter différentes applications, la représentation des connaissances sous forme d'ontologie a atteint ses limites à cause de sa restriction au point de vue d'un domaine spécifique sous forme de silo isolé, et l'alignement d'ontologies basée sur la concrétisation d'un consensus est difficile. Les graphes de connaissances (Ehrlinger, Wöß, 2016 ; Hogan *et al*., 2021 ; Gutiérrez, Sequeda, 2021) ont émergés comme une réponse flexible à ce besoin, en se concentrant sur l'intégration de données et connaissances hétérogènes (Li *et al*., 2021). Les graphes de connaissances ont également l'avantage de pouvoir utiliser différents modèles de graphe, tels que le graphe étiqueté ou le graphe de propriétés (Angles, 2018), et ainsi avoir accès à différents niveaux d'expressivité. Plus le modèle permet de réaliser des constructions variées à partir des éléments qui le composent, plus il est expressif.

Toutefois, l'expressivité des modèles de graphe actuels reste limitée, et certaines relations complexes, utiles lors de la modélisation des connaissances, sont difficilement représentables. Par exemple, on peut vouloir regrouper un sous-ensemble du graphe en un nouvel élément et lier cet élément au reste du graphe ou encore lui ajouter des métadonnées telles qu'une date de validité.

Dans cet article, nous proposons d'étendre le modèle de graphe de propriétés, en ajoutant un mécanisme d'abstraction s'inspirant de la réification RDF (Orlandi *et al*., 2021) afin d'améliorer son expressivité. De cette manière, il devient possible de représenter différents niveaux d'abstraction au sein d'un graphe de connaissance, d'ajouter des propriétés et des étiquettes aux niveaux d'abstraction et de définir des liens entre ces différents niveaux d'abstraction ou entre un élément du graphe et un niveau d'abstraction.

La suite de l'article est organisée de la manière suivante : la section 2 présente une étude de l'expressivité des différents modèles de graphes utilisés pour représenter des graphes de connaissances, la section 3 détaille les différentes approches existantes visant à améliorer l'expressivité des graphes de connaissances. Au niveau de la section 4, nous proposons une définition formelle de notre modèle. Dans la section 5, nous démontrons l'utilité de notre modèle sur des exemples dans le cadre du projet EASING qui a pour objectif de fournir des solutions d'hébergement à des personnes handicapées. Enfin, la section 6 conclut l'article et présente les orientations de nos travaux futurs.

## 2. Évolution de l'expressivité des modèles de graphes de connaissances

L'expressivité d'un graphe de connaissances dépend des opérateurs de construction supportés par le modèle de graphe qu'il utilise, permettant de décrire les entités et les relations. Dans cette section, nous explorerons l'expressivité de différents modèles des graphes.

En se basant sur la théorie des graphes, les modèles de graphe comportent un ensemble non vide de nœuds $V$ représentant les entités et un ensemble d'arêtes $E$ reliant des nœuds de $V$.

Le modèle du **graphe non orienté** est le modèle de graphe le plus basique en termes d'expressivité. Il est défini par le triplet $(V, E)$, où $E \subseteq V \times V$. Ainsi, $E$ représente les relations bidirectionnelles ou réciproques comme par exemple la relation "AMI_DE" mais il est incapable de représenter les relations unidirectionnelles comme par exemple la relation "PARENT_DE". Afin d'améliorer l'expressivité de ce modèle, il est nécessaire de rendre les arêtes directionnelles, comme il est possible de faire avec le modèle du graphe orienté.

Un **graphe orienté** est défini par le triplet $(V, E, \rho)$, où $\rho : E \rightarrow V \times V$ est une fonction totale qui retourne le couple de nœuds associés par une arête. Par exemple, $\rho(e) = (v_1, v_2)$ indique que l'arête $e \in E$ est une arête orientée de $v_1$ à $v_2$. Cependant, s'il s'agit d'un multigraphe, c'est-à-dire s'il existe plusieurs arêtes entre les mêmes nœuds source et destination, il est difficile de différencier la signification de chaque arête. Afin d'améliorer l'expressivité du modèle de graphe orienté, le modèle du graphe orienté étiqueté peut être utilisé.

Un **graphe étiqueté** permet d'associer une étiquette à chaque nœud ou chaque arête. Un graphe étiqueté est défini comme un quadruplet $(V, E, \rho, \lambda)$, où $\lambda : (E \cup V) \rightarrow \mathcal{L}$ est une fonction totale, avec $\mathcal{L}$ étant un ensemble d'étiquettes. Toutefois, cette représentation ne permet pas d'ajouter de métadonnées sur les liens ou les nœuds, contrairement au graphe de propriétés.

Un **graphe de propriétés** (Rodriguez, Neubauer, 2010 ; Angles, 2018) est défini comme un quintuplet $(V, E, \rho, \lambda, \sigma)$, où $\sigma : (V \cup E) \times Prop \rightarrow Val$ est une fonction partielle, dont $Prop$ est un ensemble fini des propriétés et $Val$ est un ensemble de valeurs. Si $v \in V$ (resp., $e \in E$), $p \in Prop$, et $\sigma(v, p) = s$ (resp., $\sigma(e, p) = s$), alors $s$ est la valeur de la propriété $p$ du nœud $v$ (resp., de l'arête $e$). Par rapport au modèle étiqueté, la fonction $\lambda : (V \cup E) \rightarrow 2^{\mathcal{L}}$, permet de définir un ensemble d'étiquettes pour chaque nœud et arête. Cependant, ce modèle de graphe, tout comme les modèles précédents, ne peut représenter que des relations binaires. En revanche, les hypergraphes dépassent cette limite et permettent de modéliser des relations n-aires.

Un **hypergraphe** (Berge, 1972) est une généralisation d'un graphe, capable de connecter plus de deux nœuds. Un hypergraphe $H = (V, E)$ est défini comme une famille d'hyper-arêtes $E$, où chaque hyper-arête est un sous-ensemble non vide de $V$, $V$ étant un ensemble fini de nœuds. Le modèle d'hypergraphe a été utilisé pour repré-

senter les relations entre plusieurs entités, telles les nœuds de type diplôme, personne et université, liés par une hyper-arête représentant l'obtention du diplôme.

TABLEAU 1. *Tableau comparatif des capacités de représentation des différents modèles de graphe*

| Modèle de graphe | Capacités de représentation | | | | | | |
|---|---|---|---|---|---|---|---|
| | Arête orientée | Arête étiquetée | Nœud étiqueté | Propriété d'arête | Propriété du nœud | Relation binaire | Relation n-aire |
| Graphe non-orienté | | | | | | ✓ | |
| Graphe orienté | ✓ | | | | | ✓ | |
| Graphe étiqueté | ✓ | ✓ | ✓ | | | ✓ | |
| Graphe de propriétés | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Hypergraphe | | | | | | ✓ | ✓ |

L'étude des modèles de graphe, synthétisée dans le tableau 1, montre d'une part que le modèle du graphe de propriétés généralise la plupart des modèles de graphe, et d'autre part que l'hypergraphe est le seul modèle qui puisse représenter les relations n-aires, sans toutefois permettre d'ajouter des étiquettes ou des propriétés. De plus, certains types de relations complexes, tels qu'un lien entre une sous partie d'un graphe et un de ses nœuds, ne peuvent être représentés par aucun de ces modèles. Dans la section suivante, nous explorons les travaux connexes qui ont essayé d'améliorer l'expressivité des modèles pour les graphes de connaissances.

## 3. Travaux connexes

RDF est un modèle pour échanger des données et des connaissances sur le Web. RDF est standardisé par le W3C et se base sur un modèle de graphe étiqueté. RDF représente des faits sous la forme de triplet (sujet, prédicat, objet) liant deux nœuds (le sujet et l'objet) par une arête orientée et étiquetée (le prédicat). Il s'agit de la seule construction du modèle. Cela implique que les faits et les propriétés d'objet sont représentés au même niveau, et qu'il est impossible de représenter des relations complexes. Par exemple, le fait "Le bâtiment 07 est accessible aux personnes aveugles depuis octobre 2023" pourrait être représenté en syntaxe Turtle avec deux triplets qui ne peuvent pas être liés :

```
:Building07 :AccessibleTo :Blind.
:Fact :since :October2023.
```

Par conséquent, il est nécessaire d'avoir un mécanisme qui exprime les métadonnées des triplets. La réification (Orlandi *et al.*, 2021) permet d'abstraire un triplet RDF pour

lui ajouter des métadonnées. Quatre types de réification RDF ont été proposés : la réification standard, la réification via les graphes nommés, la réification via la propriété singleton et la réification RDF*.

La **réification standard** (Manola, Miller, 2004), connue sous le terme RDF Primer, fait référence au processus de représentation d'un triplet RDF en tant que nouvelle ressource, qui est une instance de la classe `Statement`, avec les principales propriétés `rdf:subject`, `rdf:predicate` et `rdf:object`. Avec la réification standard, le fait "Le bâtiment 07 est accessible aux personnes aveugles depuis octobre 2023" peut être représenté comme suit:

```
_:x rdf:type rdf:Statement.
_:x rdf:subject :Building07.
_:x rdf:predicate :AccessibleTo.
_:x rdf:object :Blind.
_:x  :since :October2023.
```

Il est important de noter que la nouvelle ressource de type `rdf:Statement` est représentée, dans un graphe RDF, sous la forme d'un nœud anonyme (ou encore *blank node*), sans IRI. Chen *et al*. (2012) ont expliqué l'utilité des *blank nodes* dans un graphe RDF notamment pour relier un nœud abstrait à des métadonnées.

La **réification basée sur les graphes nommés** (Carroll *et al*., 2005) définit un graphe nommé $G$ comme un sous-graphe associé à un identifiant unique. Il est représenté sous la forme d'une paire $(G, ID)$, où $ID$ est l'identifiant du graphe nommé $G$. La notion de graphe nommé a été par la suite standardisée par le W3C (Ivan, 2010). Cette approche de réification utilise le graphe nommé comme un contexte dans lequel les triplets RDF réifiés sont regroupés. Ce mécanisme peut être utilisé pour associer un triplet principal (comme `:Building07 :AccessibleTo :Blind.`) et les triplets qui le décrivent (comme `:Fact :since :October2023.`). L'exemple précédent peut être représenté comme suit :

```
:Building07Info {
:Building07 :AccessibleTo :Blind.
:Fact :since :October2023.}
```

En se basant sur l'idée des graphes nommés, Stoermer *et al*. (2006) ont proposé d'améliorer l'expressivité du graphe RDF en permettant d'associer à un triplet un ou plusieurs contextes représentés au moyen de graphes nommés. Les auteurs ont également modélisé les relations entre les contextes par le triplet $(c_x, R, c_y)$ où $c_x$ et $c_y$ sont les IRI de deux graphes nommés représentant deux contextes différents et $R$ est l'IRI de la relation entre eux. D'autres approches se sont intéressées à l'abstraction des graphes orientés étiquetés. Poulovassilis et Levene (1994) ont proposé l'*hypernode graph*, qui permet de représenter un sous-graphe d'un graphe orienté étiqueté sous forme d'un nœud. Les sous-graphes étant des nœuds ils peuvent alors être reliés entre eux. Les *nested graphs* (Chein *et al*., 1998) proposent également un mécanisme d'abstraction, dans lequel un sous-graphe correspondant à un certain contexte peut être représenté par un nœud pouvant être lié aux autres nœuds du graphe.

La **réification de propriété singleton** (Nguyen *et al*., 2014) ajoute un prédicat unique au prédicat original d'un triplet RDF auquel il est possible d'ajouter des métadonnées grâce à des triplets supplémentaires liés au prédicat unique considéré en tant que sujet. Il s'agit d'une approche alternative à la réification standard principalement basée sur les prédicats. Par exemple, en utilisant la propriété singleton, le fait précédent devient :

```
:Building07 :AccessibleTo#1 :Blind.
:AccessibleTo#1 rdf:singletonPropertyOf :AccessibleTo.
:AccessibleTo#1 :since :October2023.
```

Avec la réification par propriété singleton, le prédicat `AccessibleTo` du triplet original est étendu à un prédicat unique `AccessibleTo#1` qui est ensuite utilisé comme sujet pour ajouter des faits supplémentaires. Les travaux de Rosso *et al*. (2020) se sont inspirés de cette technique d'ajout des métadonnées sur les arêtes en proposant d'accorder à chaque arête le couple *(qualifier, value)* représentant le nom de la métadonnée et sa valeur. Ce modèle se rapproche alors de celui du graphe de propriétés.

La **réification RDF*** (Hartig, 2017) est une simplification syntaxique visant à simplifier la réification standard en évitant de créer de nouvelles ressources. Selon RDF*, notre exemple précédent est encodé à l'aide d'un seul triplet RDF* :

```
<<:Building07 :AccessibleTo :Blind>> :since :October2023.
```

Ce triplet RDF* imbriqué est appelé le triplet de métadonnées dont le sujet est encadré par « et ». Kovacevic *et al*. (2022) se sont basés sur la technique RDF* pour annoter les triplets de données avec des métadonnées temporelles. Ils proposent d'ajouter deux attributs au triplet réifié : `valid_from`, désignant la date de création du `timestamp` et `valid_until` désignant la date d'expiration, les deux étant liés à un *timestamp*. Xiong *et al*. (2023) ont proposé les *nested facts*, consistant à ajouter des liens entre des liens. Cette approche peut être vue comme une réification RDF* entre deux triplets.

Après avoir étudié les travaux de la littérature, on peut distinguer deux familles d'approches distinctes visant à améliorer l'expressivité des modèles pour les graphes de connaissances. En premier lieu, il existe des approches qui se rapprochent du modèle graphe de propriétés, qui permet d'englober une partie des méthodes de réification grâce aux propriétés des arêtes. En deuxième lieu, on trouve des approches qui s'apparentent aux techniques de la reification RDF, que l'on peut diviser en deux catégories : 1) la manipulation de triplets, comprenant la réification standard, RDF*, et celle de la propriété singleton, et 2) le regroupement par sous-graphe, représentée par le graphe nommé. Chaque catégorie présente ses propres limites : la catégorie de manipulation de triplets ne peut réifier qu'un seul triplet RDF (soit deux nœuds connectés), tandis que la catégorie de sous-graphe ne dispose pas d'un mécanisme d'ajout de métadonnées sur le sous-graphe lui même. Certaines approches permettent de créer des liens

entre des nœuds pouvant représenter un sous-graphe abstrait, mais ne permettent toutefois pas de relier un élément du sous-graphe avec le reste du graphe. De cette étude, deux constats peuvent être faits : 1) le modèle de graphe de propriétés permet à la fois d'abstraire la plupart des modèles de graphe ainsi que plusieurs techniques de réification, et 2) les relations complexes ne peuvent pas facilement être représentées.

## 4. Approche proposée

Pour améliorer l'expressivité du graphe de propriétés, nous nous inspirons du principe de réification afin de proposer un modèle permettant de représenter un sous-graphe (un ensemble de nœuds, arêtes, propriétés et étiquettes) sous la forme d'un nœud, agissant lui-même comme un nœud standard du graphe. Ce mécanisme peut éventuellement être récursif et ainsi permettre de modéliser différents niveaux d'abstraction. Dans la suite de la section nous expliquons les principes du modèle puis nous proposons une définition formelle.

### *4.1. Principes fondamentaux du modèle proposé*

Tout d'abord, les propriétés de données et les propriétés d'objets sont toutes deux représentées par une arête dans RDF, par contre, dans un graphe de propriétés, une propriété de données devient une propriété d'un nœud. Par conséquent, les propriétés de nœuds doivent être considérées pour conserver au moins le même niveau d'expressivité que les techniques de réification manipulant les triplets. Les propriétés d'arêtes, quant à elles, doivent aussi être considérées pour exploiter au maximum le modèle du graphe de propriétés. De plus, dans un graphe de propriétés, chaque nœud et arête peut avoir plusieurs étiquettes, plutôt qu'une seule comme c'est le cas dans un graphe RDF. Il faut donc pouvoir définir un sous-ensemble des étiquettes pour un nœud ou une arête. Les relations complexes, entre des sous-graphes, doivent également pouvoir être représentées afin d'augmenter l'expressivité du modèle.

### *4.2. Définition formelle du modèle proposé*

À partir des principes définis dans la sous-section précédente, nous proposons un modèle pouvant représenter un ensemble de nœuds et potentiellement d'arêtes (les nœuds peuvent être partiellement/totalement connectés ou même déconnectés) ainsi que certaines de leurs étiquettes et propriétés pour obtenir un sous-graphe. Ce sous-graphe prend la forme d'une nœud abstrait, pouvant avoir des étiquettes, des propriétés et des liens vers d'autres nœuds standards ou abstraits.

Le modèle proposé étend le modèle de graphe de propriétés de la manière suivante :

$$G = (V, E, R, \rho, \lambda, \sigma, \alpha)$$

où :

– $R \subseteq V$, $R$ est un sous-ensemble de $V$ contenant les nœuds réifiés;

– $\alpha : R \rightarrow SG$ est une fonction totale qui fait correspondre les nœuds réifiés à leur sous-graphe. Si $r \in R$ et $sg \in SG$, alors $\alpha(r) = sg$ où $sg$ est le sous-graphe du nœud réifié $r$. $SG$ est défini par un ensemble fini de tuples $sg$ de la forme suivante :

$$sg = (V_r, E_r, R_r, \rho_r, \lambda_r, \sigma_r, \alpha_r)$$

où : $V_r \subseteq V$, $E_r \subseteq E$, $R_r \subseteq R$ avec les fonctions :

- $\rho_r = \rho|_{E_r}$
- $\lambda_r = \lambda|_{V_r \cup E_r}$ c'est-à-dire $\lambda_r : (V_r \cup E_r) \rightarrow 2^{\mathcal{L}_r}$
- $\sigma_r = \sigma|_{(V_r \cup E_r) \times Prop_r}$
- $\alpha_r = \alpha|_{R_r}$

Les fonctions $\rho_r, \lambda_r, \sigma_r$ et $\alpha_r$ sont les restrictions des fonctions $\rho, \lambda, \sigma$ et $\alpha$ qui associent respectivement à une arête orientée ses nœuds, à un nœud ou une arrête un ensemble d'étiquettes, à un nœud ou une arrête ses propriétés et leur valeur, et à un nœud réifié son sous-graphe. $Prop_r$ est un ensemble fini de propriétés sélectionnées du sous-graphe et $\mathcal{L}_r$ est un ensemble fini d'étiquettes sélectionnées, donc $Prop_r \subseteq Prop$ et $\mathcal{L}_r \subseteq \mathcal{L}$.

Le modèle ainsi défini permet alors les constructions suivantes :

– **l'abstraction d'un sous-graphe sous forme d'un nœud réifié** au moyen de la fonction $\sigma$. Un nœud réifié se compose alors d'un ensemble de nœuds, d'arêtes ainsi que de certaines de leurs étiquettes et propriétés ;

– **l'abstraction récursive** car $R \subset V$. Ainsi, un nœud réifié peut contenir d'autres nœuds réifiés et définir de cette manière des niveaux d'abstraction ;

– **l'ajout de métadonnées sur un sous-graphe** puisqu'un nœud réifié est un élément de $V$, donc il peut avoir à la fois des étiquettes et des propriétés. Il peut également avoir des liens avec d'autres nœuds standards ou réifiés du graphe.

### *4.3. Spécification du nœud réifié*

Pour construire un nœud réifié, il faut tout d'abord spécifier les éléments concernés par la réification. Comme indiqué dans la définition du modèle, il est possible de réifier les nœuds, les arêtes, ainsi que certaines de leurs étiquettes et propriétés. Donc, la fonction de construction du nœud réifié doit accepter en paramètres ces éléments individuellement ou toute combinaison d'entre eux. Pour ce faire, nous proposons la fonction suivante :

$$\beta(V_r, E_r, \lambda_r, \sigma_r) = r$$

où:

– $V_r$ est l'ensemble des nœuds sélectionnés

– $E_r$ est l'ensemble des arêtes sélectionnées

- $\lambda_r : (V_r \cup E_r) \rightarrow 2^{\mathcal{L}_r}$ se restreint sur les étiquettes sélectionnées
- $\sigma_r : (V_r \cup E_r) \times Prop_r \rightarrow Val$ se restreint sur les propriétés sélectionnées
- $\rho_r = \rho|_{E_r}$ est une restriction de la fonction $\rho$ à $E_r$

Pour obtenir $R_r$ tel qu'il est indiqué dans le modèle, nous nous appuyons sur $R$ du graphe global, car les nœuds réifiés d'un sous-graphe sont également des nœuds réifiés dans le graphe contenant le sous-graphe. Par conséquent, $R_r = \{v | v \in R \cap V_r\}$.

Étant donné que $\lambda_r$ et $\sigma_r$ ont leurs domaines restreints selon les nœuds et les arêtes sélectionnés, cela garantit que les étiquettes et propriétés sélectionnées appartiennent à un nœud ou une arête existant dans le sous-graphe de la réification.

## 5. Validation préliminaire

Dans cette section, nous identifions des cas d'utilisation réels qui illustrent la nécessité de représenter des relations complexes dans un graphe de connaissances. Nous montrons comment notre approche répond aux exigences de chaque cas d'utilisation.

Les cas d'utilisation abordés dans cette section sont dérivés de l'un des *Work Packages* du projet EASING (mobility of pErsons And acceSsible housINg). Il s'agit d'un projet français visant à aider les personnes handicapées à trouver un logement adapté à leurs besoins. Nous nous concentrons sur la vérification de la conformité des bâtiments par rapport au Code de la Construction et d'Habitation (CCH) régissant l'accessibilité pour les personnes handicapées.

Les éléments du bâtiment sont représentées à l'aide de l'ontologie IfcOWL [1] utilisée comme un schéma du graphe de connaissances utilisant le modèle du graphe de propriétés. Pour le CCH, nous nous référons aux articles du décret français du 20 avril 2017 [2] régissant l'accessibilité des établissement publics pour les personnes handicapées.

Certains termes du CCH ne correspondent pas directement aux entités du schéma du graphe de connaissances qui représente le bâtiment. Par conséquent, la mise en place d'un mécanisme de vérification de la conformité des bâtiments nécessite de représenter les termes du code de construction en fonction des éléments du bâtiment.

Nous utilisons trois cas d'utilisation différents pour montrer les exigences particulières en terme de représentation des connaissances et pour démontrer comment notre modèle peut fournir une solution appropriée. Au niveau du 1er cas d'utilisation "Représentation du terme : circulation horizontale", nous appliquons la réification sur des nœuds connectés avec leurs étiquettes et propriétés. Pour le 2e cas d'utilisation "Représentation du terme : cheminement extérieur" nous appliquons la réification récursive dont un nœud réifié contient un autre nœud réifié. Pour le 3e cas d'utilisation "Repré-

1. https://github.com/buildingSMART/ifcOWL/blob/master/IFC2X3_Final.owl
2. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000034485459/

sentation du terme : circulation verticale" nous appliquons la réification des nœuds non reliés ainsi que de leurs étiquettes en ajoutant un lien entre des nœuds réifiés.

### *5.1. Représentation du terme "Circulation horizontale"*

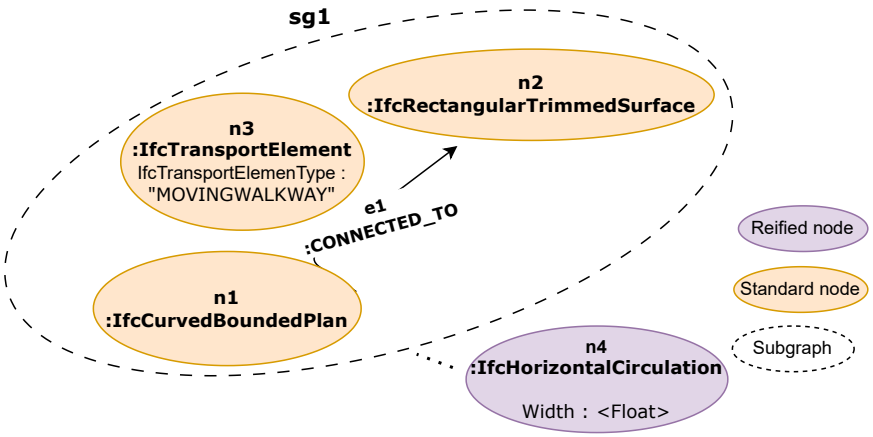



FIGURE 1. *La représentation du terme de circulation horizontale en utilisant la réification*

Le sixième article du CCH indique : "Les circulations horizontales doivent avoir une largeur minimale de 1,40 mètre". Le terme utilisé dans le CCH est "circulation horizontale". Dans la représentation IfcOWL, les éléments du bâtiment constituant ce terme sont les suivants : un plan borné courbé (IfcCurvedBoundedPlan) connecté à une surface rectangulaire rognée (IfcRectangularTrimmedSurface), et un élément de transport (IfcTransportElement) qui doit être une passerelle. Nous devons réifier les éléments ci-dessus en un nœud abstrait étiqueté IfcHorizontalCirculation.

Pour ce faire, nous créons un nœud réifié IfcHorizontalCirculation via la réification des trois nœuds $n_1$, $n_2$, $n_3$, de l'arête $e_1$, des étiquettes "IfcCurvedBoundedPlan" de $n_1$,"IfcRectangularTrimmedSurface" de $n_2$, "IfcTransportElement" de $n_3$ et de l'étiquette "CONNECTED_TO" de $e_1$, ainsi que de la propriété "IfcTransportElementType" égale à "MOVINGWALKWAY" de $n_3$ (voir la figure 1). En appliquant la définition de notre modèle, le sous-graphe $sg_1$ qui constitue le nœud réifié IfcHorizontalCirculation est le suivant:

$$sg_1 = (V_r, E_r, R_r, \rho_r, \lambda_r, \sigma_r, \alpha_r)$$

où:

– $V_r = \{n_1, n_2, n_3\}$, $E_r = \{e_1\}$, $R_r = \emptyset$

- $\rho_r(e_1) = (n_1, n_2)$
- $\lambda_r(n_1) = "IfcCurvedBoundedPlan"$
- $\lambda_r(n_2) = "IfcRectangularTrimmedSurface"$
- $\lambda_r(n_3) = "IfcTransportElement"$
- $\lambda_r(e_1) = "CONNECTED\_TO"$
- $\sigma_r(n_3, "IfcTransportElementType") = "MOVINGWALKWAY"$

$$\beta(V_r, E_r, \lambda_r, \sigma_r) = n_4$$

### *5.2. Représentation du terme "Cheminement extérieur"*

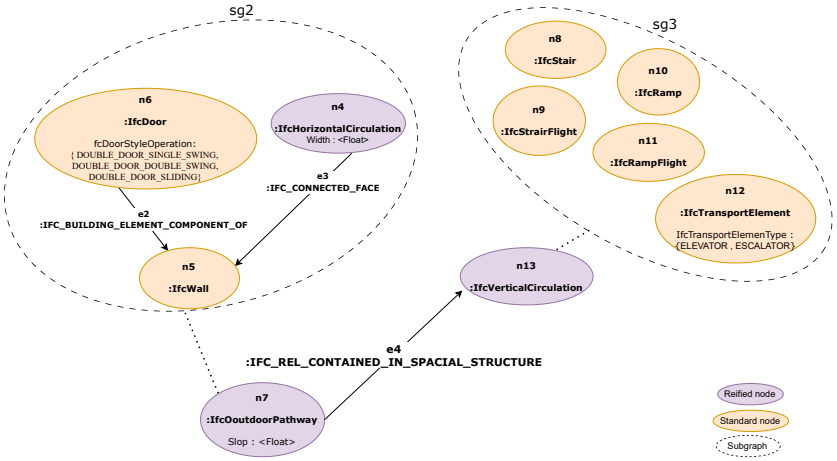



FIGURE 2. *Représentation des termes Cheminement Extérieur et Circulation Verticale en utilisant la réification*

Le deuxième article du CCH indique : "Si un cheminement extérieur a une pente de 5% ou moins, il doit disposer d'un équipement de circulation verticale". Le premier terme du code est "Cheminement Extérieur" ou encore *Outdoor Pathway*. Un chemin extérieur est un cheminement horizontal se terminant par une porte. Ainsi, les nœuds constituant ce terme sont étiquetés IfcDoor et IfcHorizontalCirculation. Pour la représentation du terme "Cheminement Extérieur", le nœud réifié IfcHorizontalCirculation qui a été précédemment créé servira à représenter le terme "Cheminement Extérieur".

Ainsi, nous sélectionnons le nœud réifié $n_4$ avec son étiquette IfcHorizontalCirculation, le nœud standard $n_5$ avec son étiquette IfcWall, le nœud standard $n_6$ avec son étiquette IfcDoor avec sa propriété IfcDoorStyleOperation, l'arête $e_2$ avec son étiquette IFC_BUILDING_ELEMENT_COMPONENT_OF et l'arête $e_3$ avec son étiquette IFC_CONNECTED_FACE (voir la partie gauche de la figure 2).

Nous pouvons noter que le nœud réifié IfcOutdoorPathway contient un autre nœud réifié IfcHorizontalCirculation. Il s'agit d'une réification récursive qui permet de représenter plusieurs niveaux d'abstraction. En appliquant la définition de notre modèle, le sous-graphe $sg_2$ qui constitue le nœud réifié IfcOutdoorPathway est le suivant:

$$sg_2 = (V_r, E_r, R_r, \rho_r, \lambda_r, \sigma_r, \alpha_r)$$

où:

- $V_r = \{n_4, n_5, n_6\}$, $E_r = \{e_2, e_3\}$, $R_r = \{n_4\}$
- $\rho_r(e_2) = (n_5, n_6)$
- $\rho_r(e_3) = (n_4, n_5)$
- $\lambda_r(n_4) = "IfcHorizontalCirculation"$
- $\lambda_r(n_5) = "IfcWall"$
- $\lambda_r(n_6) = "IfcDoor"$
- $\lambda_r(e_2) = "IFC\_BUILDING\_ELEMENT\_COMPONENT\_OF"$
- $\lambda_r(e_3) = "IFC\_CONNECTED\_FACE"$
- $\sigma_r(n_6, "IfcDoorStyleOperation") = \{"DOUBLE\_DOOR\_SINGLE\_SWING", "DOUBLE\_DOOR\_DOUBLE\_SWING", "DOUBLE\_DOOR\_SLIDING"\}$
- $\alpha_r(n_4) = sg_1$

$$\beta(V_r, E_r, \lambda_r, \sigma_r) = n_7$$

### *5.3. Représentation du terme "Circulation Verticale"*

Dans son deuxième, septième et seizième article, le CCH utilise le terme "Circulation Verticale". Les nœuds constituant ce terme sont : IfcStair, IfcStairFlight, IfcRamp, IfcRampFlight et IfcTransportElement. Le terme "Circulation Verticale" ou *Vertical Circulation* représente des nœuds qui sont totalement déconnectés (aucune arête entre eux).

Ainsi, pour créer un nœud réifié étiqueté IfcVerticalCirculation, nous sélectionnons les nœuds $n_8$, $n_9$, $n_{10}$ et $n_{11}$ avec leur étiquette respective, ainsi que le nœud $n_{12}$ avec son étiquette IfcTransportElement et sa propriété IfcTransportElementType (voir la partie droite de la figure 2). En appliquant la définition de notre modèle, le sous-graphe $sg_3$ qui constitue le nœud réifié IfcVerticalCirculation est le suivant:

$$sg_3 = (V_r, E_r, R_r, \rho_r, \lambda_r, \sigma_r, \alpha_r)$$

où:

- $V_r = \{n_8, n_9, n_{10}, n_{11}, n_{12}\}$, $E_r = \emptyset$, $R_r = \emptyset$
- $\lambda_r(n_8) = "IfcStair"$
- $\lambda_r(n_9) = "IfcStairFlight"$
- $\lambda_r(n_{10}) = "IfcRamp"$

- $\lambda_r(n_{11}) = "IfcRampFlight"$
- $\lambda_r(n_{12}) = "IfcTransportElement"$
- $\sigma_r(n_{12}, "IfcTransportElementType") = \{"ELEVATOR", "ESCALATOR"\}$

$$\beta(V_r, E_r, \lambda_r, \sigma_r) = n_{13}$$

À travers les cas d'utilisation traités ci-dessus, nous avons démontré comment notre approche a contribué à améliorer l'expressivité du graphe de propriétés en représentant des relations complexes à l'aide de l'abstraction des nœuds, des arêtes, de leurs propriétés et étiquettes ou bien en ajoutant des niveaux d'abstraction avec réification récursive.

## 6. Conclusion et perspectives

Afin de pouvoir augmenter l'expressivité des graphes de connaissances et représenter des relations complexes, nous avons proposé un modèle basé sur le graphe de propriétés qui permet d'abstraire un sous-graphe sous forme de nœud se comportant comme un nœud standard du graphe, afin de pouvoir lui ajouter des liens, des propriétés et des étiquettes. Grâce à la récursivité de la définition, il devient alors possible de définir plusieurs niveaux d'abstraction avec leurs métadonnées. Nous avons également illustré l'utilité du modèle proposé à travers des exemples concrets dans le cadre du projet EASING.

Nos travaux futurs se concentrent sur deux axes principaux. Premièrement, nous prévoyons de passer à l'implémentation technique de notre approche. Deuxièmement, nous prévoyons d'adapter les langages de requête de graphe de propriétés actuels à notre modèle. Cette adaptation est essentielle pour permettre une gestion efficace de l'abstraction à plusieurs niveaux lors de l'interrogation des graphes de connaissances. Par exemple, les requêtes devront être capables d'assurer l'agrégation des propriétés, notamment lorsqu'il s'agit de calculer une propriété à partir des propriétés des éléments constitutifs d'un nœud abstrait. Ces deux axes représentent une étape cruciale pour exploiter pleinement le potentiel de notre modèle dans des contextes pratiques et pour faciliter son intégration dans les applications et les systèmes informatiques existants.

## Bibliographie

Angles R. (2018). The property graph database model. In *Amw*.

Berge C. (1972). Graphes et hypergraphes. *Dunod*.

Carroll J. J., Bizer C., Hayes P., Stickler P. (2005). Named graphs. *Journal of Web Semantics*, vol. 3, nº 4, p. 247–267.

Chein M., Mugnier M.-L., Simonet G. (1998). Nested graphs: A graph-based knowledge representation model with fol semantics. In *Kr*, p. 524–535.

Chen L., Zhang H., Chen Y., Guo W. (2012). Blank nodes in rdf. *J. Softw.*, vol. 7, nº 9, p. 1993–1999.

Ehrlinger L., Wöß W. (2016). Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, vol. 48, nº 1-4, p. 2.

Gutiérrez C., Sequeda J. F. (2021). Knowledge graphs. *Communications of the ACM*, vol. 64, nº 3, p. 96–104.

Hartig O. (2017). Foundations of RDF* and SPARQL*:(an alternative approach to statement-level metadata in RDF). In *Amw 2017 11th alberto mendelzon international workshop on foundations of data management and the web, montevideo, uruguay, june 7-9, 2017.*, vol. 1912.

Hogan A., Blomqvist E., Cochez M., d'Amato C., Melo G. D., Gutierrez C. *et al.* (2021). Knowledge graphs. *ACM Computing Surveys (CSUR)*, vol. 54, nº 4, p. 1–37.

Ivan H. (2010). *Rdf graph literals and named graphs.* Consulté sur https://www.w3.org/2009/07/NamedGraph.html

Kovacevic F., Ekaputra F. J., Miksa T., Rauber A. (2022). Starvers-versioning and timestamping rdf data by means of rdf*-an approach based on annotated triples.

Li X., Lyu M., Wang Z., Chen C.-H., Zheng P. (2021). Exploiting knowledge graphs in industrial products and services: a survey of key aspects, challenges, and future perspectives. *Computers in Industry*, vol. 129, p. 103449.

Manola F., Miller E. (2004). *Rdf reification.* Consulté sur {https://www.w3.org/TR/rdf-primer/\#reification}

Nguyen V., Bodenreider O., Sheth A. (2014). Don't like RDF reification? making statements about statements using singleton property. In *Proceedings of the 23rd international conference on world wide web*, p. 759–770.

Orlandi F., Graux D., O'Sullivan D. (2021). Benchmarking RDF metadata representations: Reification, singleton property and RDF. In *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, p. 233–240.

Poulovassilis A., Levene M. (1994). A nested-graph model for the representation and manipulation of complex objects. *ACM Transactions on Information Systems (TOIS)*, vol. 12, nº 1, p. 35–68.

Rodriguez M. A., Neubauer P. (2010). Constructions from dots and lines. *arXiv preprint arXiv:1006.2361*.

Rosso P., Yang D., Cudré-Mauroux P. (2020). Beyond triplets: hyper-relational knowledge graph embedding for link prediction. In *Proceedings of the web conference 2020*, p. 1885–1896.

Stoermer H., Palmisano I., Redavid D., Iannone L., Bouquet P., Semeraro G. *et al.* (2006). rdf and contexts: Use of sparql and named graphs to achieve contextualization. In *Proc. of 2006 jena user conference.*

Xiong B., Nayyeri M., Luo L., Wang Z., Pan S., Staab S. (2023). Neste: Modeling nested relational structures for knowledge graph reasoning. *arXiv preprint arXiv:2312.09219*.

# Exploration des pratiques de régulation des IA génératives par le protocole d'exclusion des robots

**Robert Viseur[1], Landelin Delcoucq[2]**

*1. Service TIC, FWEG, UMONS*
*17 place Warocqué, B-7000 Mons, Belgique*
*robert.viseur@umons.ac.be*

*2. Service MIT, FPMs, UMONS*
*9 rue de Houdain, B-7000 Mons, Belgique*
*landelin.delcoucq@umons.ac.be*

*RÉSUMÉ. L'année 2023 fut notamment celle de l'essor des IA génératives capables de produire des images (Stable Diffusion, Midjourney…) ou des textes (ChatGPT, Bard…) originaux. Ces nouveaux outils ont amené leur lot de polémiques. Parmi celle-ci, la question des droits d'auteurs des contenus utilisés pour l'entraînement de ces modèle a rapidement touché les scènes médiatiques puis judiciaires. Dans cette recherche exploratoire, nous avons utilisé un robot d'exploration pour analyser les fichiers « robots.txt » de plusieurs ensemble de sites web incluant le Top 100 Alexa, des sites de presse en ligne et des sites d'éditeurs scientifiques. L'objectif était d'analyser le recours à cette norme technique, soit le protocole d'exclusion des robots, pour traiter cette question de la violation de la propriété intellectuelle. Nos résultats montrent une forte utilisation des mesures de blocage par les sites vivant de la publication de contenus. Ils mettent cependant en évidence certaines incohérences dans les mesures de blocage, des limitations dans le protocole d'exclusion des robots et des biais (pour lesquels une nouvelle mesure est proposée) que les politiques de blocage différenciées risquent d'introduire lors de l'entraînement des IA génératives.*

*ABSTRACT. The year 2023 saw the rise of generative AI capable of producing original images (Stable Diffusion, Midjourney...) or texts (ChatGPT, Bard...). These new tools have brought with them their share of controversy. One of these has been the issue of copyright in the content used to train these models, which has rapidly made its way into the media and legal arena. In this exploratory study, we used a crawler to analyse the "robots.txt" files of several sets of websites, including the Alexa Top 100, online press sites and scientific publishers' sites. The aim was to analyse the use of this technical standard, the robot exclusion protocol, to address the issue of intellectual property infringement. Our results show a high level of use of blocking measures by sites publishing content. However, they highlight certain inconsistencies in the blocking measures, limitations in the bot exclusion protocol and the biases (for which a new measure is proposed) that differentiated blocking policies are likely to cause when training generative AIs.*

*Mots-clés : IA, biais, LLM, ChatGPT, robots d'exploration, propriété intellectuelle.*

*KEYWORDS: AI, bias, LLM, ChatGPT, crawler, intellectual property.*

## 1. Introduction

L'intelligence artificielle peut être vue comme « *un artefact informatique construit grâce à l'intervention humaine, qui pense ou agit comme les humains, ou comme nous nous attendons à ce que les humains pensent ou agissent* » (Dignum, 2019). Elle couvre différentes approches techniques incluant le *machine learning*. Ce dernier a connu des progrès sensibles ces dernières années avec l'essor des réseaux de neurones profonds, ou *deep learning* (Heudin, 2016). Dopé par les ressources de calcul disponibles (GPU, GPU as a Service…), le *deep learning* a notamment permis d'améliorer les dispositifs de reconnaissance de forme dans les images (p. ex. réseaux de neurones convolutifs). Plus récemment, les intelligences artificielles basées sur le *deep learning* ont démontré des aptitudes à la créativité, d'abord avec les *Generative Adversial Networks* (GAN), puis avec les *Latent Diffusion Models* (LDM), pour la création d'images, et les *Large Language Models* (LLM), pour la création de textes (Goodfellow et al., 2014 ; Floridi & Chiriatti, 2020 ; Roombach et al., 2022). Ces nouveaux outils (Stable Diffusion, Midjourney, OpenAI DALL-E, OpenAI ChatGPT, Google Bard…) se sont, en moins de 2 ans, rapidement diffusés auprès du grand public. À côté de celle sur l'authenticité de cette créativité (Chomsky, 2023), une polémique relative à la propriété intellectuelle a rapidement émergé, notamment dans le secteur de la presse (McKenzie & Arvanitis, 2023). Plusieurs actions en justice sont ainsi en cours contre OpenAI[1]. Parmi les acteurs attentifs à la défense de leur propriété intellectuelle face à ces nouveaux entrants citons en particulier le New York Times (Weatherbed, 2023). En pratique, ces IA sont entraînées sur des volumes massifs de données. Ces dernières sont souvent collectées par des robots d'exploration sur le Web sans concertation préalable et explicite avec les gestionnaires des sites web. Certains jeux de données, comme LAION pour les images et Common Crawl pour le texte, sont ainsi maintenus et publiés par des organisations sans but lucratif. Ce conflit quant à la réutilisation de contenus publiés en ligne n'est pas radicalement nouveau. Il rappelle ainsi les conflits récurrents entre la presse en ligne et Google autour, notamment, de son service Google Actualités (Rebillard & Smyrnaios, 2010 ; Ouakrat, 2020 ; Galloway, 2018). Ce différend s'est réglé par un mélange de négociations individuelles, de régulation et de normes techniques incluant le protocole d'exclusion des robots (Sun et al., 2007). Ce dernier permet aux gestionnaires de sites de limiter l'activité des robots éthiques sur leurs sites. Nous proposons dans cette recherche exploratoire d'étudier comment les gestionnaires de sites web recourent au protocole d'exclusion des robots et d'en discuter l'efficacité. La suite de notre papier est découpée en quatre sections : la revue de la littérature, la présentation de la méthodologie et des données utilisées, les résultats puis leur discussion (en particulier sur le plan des risques d'introduction de biais).

## 2. Revue de la littérature

Cette section présente les intelligences artificielles génératives puis les questions de propriété intellectuelle qu'elles soulèvent. Elle se termine par une description du protocole d'exclusion des robots suivie d'une présentation des robots utilisés par les principales IA génératives.

[1] Voir https://originality.ai/blog/openai-chatgpt-lawsuit-list.

### *2.1. Intelligences artificielles génératives*

ChatGPT, développé par OpenAI, est un agent conversationnel (*chatbot*) généraliste basé sur l'architecture GPT (*Generative Pre-trained Transformer*)[2]. Il se distingue par sa faculté à interpréter des directives formulées en langage naturel, appelées « *prompts* », et à engendrer des réponses textuelles cohérentes en adéquation avec ces directives. GPT est « *un modèle linguistique autorégressif de troisième génération qui utilise l'apprentissage profond pour produire des textes semblables à ceux des humains* » (Floridi & Chiriatti, 2020). En tant que *Large Langage Model* (LLM), il est capable de « *prédire statistiquement des séquences de mots* ». Ce modèle linguistique est formé par entraînement « *sur un ensemble de données non étiquetées composé de textes, tels que Wikipédia et de nombreux autres sites, principalement en anglais, mais aussi dans d'autres langues* » (Floridi & Chiriatti, 2020). Le Common Crawl, filtré, représente ainsi 60 % des données d'entraînement de GPT-3 (Brown et al., 2020). ChatGPT est actuellement disponible en deux versions : ChatGPT (gratuit), basé sur GPT-3.5, et ChatGPT Plus (20 dollars / mois), donnant accès à GPT-4 ainsi qu'à des fonctionnalités complémentaires incluant l'accès au Web (navigation) ou l'accès à des extensions tierces (Hackett, 2023 ; OpenAI).

ChatGPT s'est distingué par son rythme de croissance extrêmement rapide. En deux mois, il atteignait ainsi les cent millions d'utilisateurs actif (Hu, 2023). D'autres agents conversationnels en mode SaaS sont cependant venus concurrencer ChatGPT. Ils incluent notamment Bard, rebaptisé Gemini début 2024, chez Google, et Claude, chez Anthropic (Singh et al., 2023). La concurrence s'est également développée du côté des LLMs. La jeune pousse Hugging Face s'est illustrée avec sa communauté éponyme dédiée à la diffusion de jeux de données et de modèles open-sources[3]. Deux modèles ont particulièrement fait l'actualité. Le premier est le modèle proposé par la jeune pousse française Mistral (Jiang et al., 2023). Le second est le modèle Llama publié par META (Touvron et al., 2023). L'utilisation locale de modèles open-sources s'est par ailleurs trouvée simplifiée par la publication de plateformes technologiques telles qu'Ollama.

### *2.2. Questions de propriété intellectuelle*

Les craintes suscitées en matière de droit d'auteur par les IA génératives portent, d'une part, sur les réponses aux *prompts*, d'autre part, sur les informations utilisées, lors de la phase d'entraînement, pour créer le modèle. Les IA génératives telles que ChatGPT se distinguent en effet par leur « *capacité à générer des textes dans n'importe quelle langue, dans n'importe quel format et sur n'importe quel sujet en quelques secondes* » (Lucchi, 2023). La question se pose donc de savoir si ces réponses sont elles-mêmes soumises au droit d'auteur. Dès lors que les États-Unis et l'Union européenne imposent la présence d'un humain en tant que créateur de l'œuvre, les productions des IA génératives (soit des machines) ne sont pas protégées par un droit d'auteur (Lucchi, 2023 ; Zirpoli, 2023). Deux cas particuliers doivent cependant être distingués. Le premier concerne la production d'une réponse trop proche des données d'entraînement. Dans ce cas, la similitude peut conduire à la reconnaissance d'une contrefaçon de l'œuvre originale. Le second concerne

[2] Voir https://help.openai.com/en/articles/6783457-what-is-chatgpt.
[3] Voir https://huggingface.co/datasets et https://huggingface.co/models.

l'existence d'une éventuelle paternité dans le cas où l'utilisateur fournit des données (« *input* ») au sein du *prompt* ou conçoit une séquence de *prompts* (« *instruction* ») particulièrement élaborée (Lucchi, 2023). La reconnaissance liée au contrôle par les instructions tend actuellement à être rejetée ; au contraire, cette reconnaissance serait avérée dès lors que l'humain apporte des modifications suffisamment créatives (Zirpoli, 2023).

La création d'un LLM passe par l'entraînement du modèle sur base d'importantes quantités de données textuelles. Ces données sont notamment collectées sur le Web. En tant qu'œuvres de l'esprit, ces documents bénéficient, par le seul acte de création, sans dépôt ni formalité, d'une protection par le droit d'auteur (Mattatia, 2017 ; Binctin, 2022). Certains de ces contenus sont par ailleurs couverts par des contrats autorisant certaines réutilisations, comme la famille des licences Creative Commons (Mattatia, 2017). Le droit d'auteur comporte également des exceptions. La question se pose ainsi de savoir si l'utilisation par les producteurs d'IA générative de ces documents a priori protégés relève du « *fair use* » (dans les pays, comme les États-Unis, où cette notion est reconnue) ou d'autres exceptions (p. ex. fouille de textes et de données) au droit d'auteur (Lucchi, 2023). Ces incertitudes juridiques, couplées à l'absence de partage de revenus, ont conduit des créateurs de contenus, d'une part, à revoir leurs conditions d'utilisation (Weatherbed, 2023), d'autre part, à attaquer en justice des producteurs d'IA génératives tels qu'OpenAI (Lucchi, 2023).

Les modèles LLM tels que GPT sont affectés par la problématique des biais. Dans ce contexte, un bias est défini comme « *la présence de déformations systématiques, d'erreurs d'attribution ou de distorsions factuelles qui ont pour effet de favoriser certains groupes ou certaines idées, de perpétuer des stéréotypes ou de rendre plus difficile l'accès à l'information* » (Ferrara, 2023). Plusieurs facteurs expliquent ce phénomène. Citons les jeux de données, les algorithmes d'apprentissage, l'annotation humaine des données et les décisions réglementaires (Ferrara, 2023). Les LLM mettent en œuvre une technique d'apprentissage auto-supervisé (Kalyan, 2023). L'auto-supervision implique que le modèle apprend à partir de données non étiquetées en générant ses propres étiquettes à partir des données elles-mêmes. La qualité du modèle généré dépend cependant de celle des données elles-mêmes (Ferrara, 2023). Il en résulte un important travail de nettoyage (Dodge et al., 2021). L'intervention humaine est utile pour assainir le jeu de données utilisé pour l'entraînement ou analyser les éventuelles défaillances lors de la génération de textes. C'est notamment ce qui explique le recours à des « *travailleurs du clic* » pour identifier les propos haineux ou violents dans ces données (Douet, 2023 ; Casilli, 2019 ; Tubaro, Casilli & Coville, 2020). En fonction des données utilisées, les biais vont prendre différentes formes. Elles incluent les biais démographiques, les biais culturels, les biais linguistiques, les biais temporels, les biais de confirmation ainsi que les biais idéologiques et politiques (Ferrara, 2023).

### *2.3. Protocole d'exclusion des robots*

La collecte de données sur le Web passe par l'utilisation de logiciels spécialisés appelés « *robots* ». Sur base d'un ensemble d'hyperliens fournis par un utilisateur, le robot va en sauvegarder le contenu, découvrir d'autres hyperliens et généralement continuer son exploration de manière récursive. Il peut être utilisé pour la création de collections, soit à des fins personnelles (p. ex. sauvegarde locale d'un site web à

l'aide de wget ou d'HTTrack), soit à des fins commerciales (p. ex. moteur de recherche : Googlebot, Bingbot...), pour l'archivage (p. ex. Internet Archive), pour la recherche personnelle ou pour la création de statistiques (p. ex. Netcraft). L'activité des robots d'exploration sur les sites web peut être régulée à l'aide de mesures actives, telles que la détection puis le blocage, ou passives, telles que le protocole d'exclusion des robots (Yang & Liao, 2010).

Le protocole d'exclusion des robots peut prendre deux formes. La première est un format de balise HTML, soit la balise META robots, permettant notamment d'indiquer si une page peut être indexée (« *index* ») ou non (« *noindex* »). La seconde est un fichier structuré nommé « *robots.txt* » placé à la racine du site web. Il documente les fichiers ou les répertoires accessibles (« *allow* ») ou non (« *disallow* ») aux robots ainsi que certains souhaits plus spécifiques comme le délai entre deux requêtes (Sun, Zhuang & Giles, 2007 ; Sun, 2008). Il présente l'allure suivante :

```
User-Agent: *
Allow: /news
User-Agent: Googlebot
Disallow: /bin
Disallow: /log
Disallow: /src
User-Agent: wget
Disallow: /
```

Ces conventions peuvent être utilisées pour exclure certains robots, soit que leur activité soit jugée nuisible aux performances du site web (p. ex. surcharge du serveur sans apport d'audience), soit que le *webmaster* bloque les robots afin d'éviter un usage non souhaité de contenus protégés par droit d'auteur (Yang & Liao, 2010). Autoriser explicitement l'accès aux robots d'un moteur de recherche pourrait donc être vu comme une licence implicite d'accéder à ces contenus (Yang & Liao, 2010). La syntaxe de ce protocole reste cependant ambiguë quant aux droits conférés, et les tentatives de clarification telles que l'*Automated Access Content Protocol* (ACAP), n'ont pas connu le succès (Sire, 2015).

### *2.4. Robots d'exploration des IA génératives*

Dans le cas de ChatGPT, plusieurs robots interviennent dans le fonctionnement de l'outil[4]. Le premier, nommé GPTBot, annoncé par OpenAI en août 2023 (David, 2023), est le robot d'exploration d'OpenAI (« *web crawling* »). Il intervient dans la collecte de données utilisées pour entraîner l'IA. Il peut être identifié par le gestionnaire du site web à partir de son *user-agent* ou de ses adresses IP. Le second, nommé ChatGPT-User, est principalement utilisée pour répondre à une requête d'un utilisateur imposant à ChatGPT d'utiliser sa fonction de navigation (« *user browsing* »). Actuellement, OpenAI gère ces robots comme un seul robot : le blocage d'un des deux entraîne donc le blocage des deux robots. À ces deux robots opérés par OpenAI doit être ajouté CCBot. Ce dernier est le robot d'exploration, basé sur le moteur de recherche open-source Nutch, utilisé par Common Crawl, une fondation à but non lucratif fournissant une copie du Web à destination des chercheurs[5]. Le Common Crawl est notamment utilisé par les chercheurs en *machine*

[4] Voir https://platform.openai.com/docs/plugins/bot.
[5] Voir https://commoncrawl.org/faq.

*learning* pour entraîner leurs modèles. Dans le cas de Google Bard, le robot d'exploration est nommé Google-Extended[6].

Ces problématiques ne sont pas totalement nouvelles. La régulation du comportement des robots d'exploration s'est en effet déjà posée à l'occasion d'un différend judiciaire qui a opposé en Belgique Copiepresse, un organisme représentant notamment la presse francophone belge, et Google, opérant le service Google Actualités (Yang & Liao, 2010). Le moteur de recherche avait notamment publié le contenu de certains articles au sein de son cache sans que cet usage n'ait été accordé (via la balise META robots « *archive* ») par les gestionnaires des sites web concernés, ce qui constitue un acte de contrefaçon au sens du droit d'auteur en Europe et a entraîné le paiement d'astreintes[7] par l'entreprise étasunienne. Ce type de contentieux a pris une dimension européenne avec la nouvelle directive européenne sur les droits voisins (Ouakrat, 2020). Les droits d'auteur ont ainsi été complétés par des droits voisins au profit des éditeurs et des agences de presse. Elle permet, contre rémunération, la reproduction et la diffusion totale ou partielle des contenus par les plates-formes (Ouakrat, 2020). Des accords ont ainsi été signés en 2023 entre Google et des organismes de gestion collective des droits, tels que DVP en France ou Corint Media en Allemagne, après une série d'accords individuels (Agence France Presse, Le Monde...).

### 3. Méthodologie et données

Afin de mieux comprendre comment les gestionnaires de sites web abordent l'arrivée sur le marché des outils d'IA générative, nous avons conçu un robot d'exploration capable de lire le contenu des fichiers « *robots.txt* » de plusieurs ensembles de sites web incluant des journaux français (27), des journaux belges francophones (10) et néerlandophones (8), des éditeurs de journaux scientifiques (16) et les sites appartenant à la dernière version publiée (2022) du Top 100 Alexa. Pour chaque ensemble nous avons analysé les directives concernant tous les robots (*) puis certains robots spécifiques incluant Bingbot, Googlebot, Googlebot-News, Wget, HTTrack, Google-Extended, CCBot, GPTbot et ChatGPT-User.

Cette sélection de robots se justifie comme suit. Bingbot et Googlebot sont les robots d'exploration liés aux moteurs de recherche Google Web Search et Bing. Leur activité ne pose généralement pas de problèmes aux yeux des gestionnaires de sites web à la recherche de visibilité dans les moteurs de recherche généralistes. Googlebot-News est le robot d'exploration spécifique à Google Actualités. L'utilisation des informations collectées par ce robot a donné lieu à des conflits judiciaires entre la presse et Google. Il permet donc de voir si les webmasters lui appliquent un traitement particulier. Wget et HTTrack sont des outils permettant l'aspiration de sites. Cette activité n'est généralement pas appréciée des gestionnaires de sites web dès lors que ces outils occasionnent une sollicitation des serveurs sans réelle contrepartie sur le plan du trafic ou des revenus. Les quatre derniers robots (soit Google-Extended, CCBot, GPTbot et ChatGPT-User) sont des robots utilisés par des producteurs de jeux de données (CCBot) ou de modèles de type LLM. C'est principalement la politique de régulation de ces robots qui nous intéresse dans le cadre de cette recherche.

---

[6] Voir https://blog.google/technology/ai/an-update-on-web-publisher-controls/.
[7] Voir https://www.copiepresse.be/judiciaire.php?classement=03.

Notre script d'analyse récupère les fichiers « *robots.txt* » des ensembles de sites web puis calcule si chaque robot est ou non cité. Ensuite, le script regarde si chaque robot fait l'objet d'un blocage complet. Le blocage complet est généralement notifié par l'instruction « *disallow* » suivie de « / ». Plusieurs exécutions ont été réalisées : 01-12-2023 et 21-01-2024. Chaque exécution sauvegarde localement les fichiers « *robots.txt* » exploités et alimente un fichier journal permettant de visualiser les résultats. Cela permet notamment de détecter d'éventuels dysfonctionnements liés à des erreurs de programmation ou à des spécificités de formatage de certains fichiers « *robots.txt* ». Ce script a par la suite été modifié pour ajouter une mesure de biais.

## 4. Analyse des résultats

Si l'on se concentre sur les 10 robots les plus fréquents (cf. Tableau 1), GPTBot ressort comme le robot d'exploration le plus cité. Les robots d'exploration des producteurs d'IA génératives font d'ailleurs l'objet d'une attention particulière des gestionnaires de sites puisque quatre d'entre eux se retrouvent dans le Top 10. Les autres robots fréquemment cités sont ceux des deux moteurs de recherche dominants, des régies publicitaires de Google, de Twitter et de l'Internet Archive. On constate également une augmentation des citations liées aux robots d'IA génératives entre les deux extractions, ce qui montre une attention croissante des gestionnaires de sites web à l'activité de ces robots.

*Tableau 1. Statistiques de mentions de robots d'exploration (01-12-2023 & 21-01-2024).*

| **Robots** | **Mentions (#)** | | **Robots** | **Mentions (#)** | |
|---|---|---|---|---|---|
| * | 137 | 137 | | | |
| gptbot | 34 | 39 | bingbot | 21 | 23 |
| twitterbot | 32 | 32 | chatgpt-user | 21 | 23 |
| googlebot | 29 | 29 | ccbot | 19 | 23 |
| mediapartners-google | 23 | 24 | google-extended | 19 | 23 |
| adsbot-google | 22 | 23 | ia_archiver | 17 | 17 |

Les gestionnaires des sites du Top 100 Alexa (cf. Tableau 2) se satisfont globalement de consignes globales fournies à l'aide du joker (« * »). Lorsqu'un robot d'exploration est cité, il s'agit le plus généralement d'un robot de moteur de recherche. GPTbot est cependant bloqué par 10,2 % des sites web (soit 29,5 % en ajoutant les blocages par défaut via le joker).

*Tableau 2. Traitement des robots par les sites du Top 100 Alexa (21-01-2024).*

| | **Robot** | **Citations** | **Allowed/Total** | **Disallowed/ Robotstxt** |
|---|---|---|---|---|
| **Site** | | 99 | | |
| **Pas de « robots.txt »** | | 11 | | |
| | * | 85 | 80,0 % | 19,3 % |
| | Bingbot | 16 | 93,8 % | 1,1 % |
| | Googlebot | 21 | 90,5 % | 2,3 % |

| | Robot | Citations | Allowed/Total | Disallowed/ Robotstxt |
|---|---|---|---|---|
| | Googlebot-News | 1 | 100,0 % | 0,0 % |
| | Wget | 1 | 0,0 % | 1,1 % |
| | HTTrack | 1 | 0,0 % | 1,1 % |
| | Google-Extended | 3 | 33,3 % | 2,3 % |
| | CCBot | 5 | 0,0 % | 5,7 % |
| | GPTBot | 9 | 0,0 % | 10,2 % |
| | ChatGPT-User | 1 | 0,0 % | 1,1 % |

L'inclusion des consignes spécifiques aux robots d'exploration des IA génératives est davantage développée chez les sites de journaux en ligne (cf. Tableau 3), qu'il s'agisse de journaux belges ou de journaux français. Nous trouvons ainsi 44,4 % des journaux francophones belges, 75 % des journaux néerlandophones belges et 40,7 % des journaux français qui procèdent au blocage de GPTBot. Ce dernier s'accompagne généralement du blocage de ChatGPT-User (mais il se révèle moins systématique). Le constat est similaire chez les sites des éditeurs scientifiques (avec une certaine tolérance à l'égard de ChatGPT-User).

*Tableau 3. Traitement des robots par les sites de presse française (21-01-2024).*

| | Robot | Citations | Allowed/Citations | Disallowed/ Robotstxt |
|---|---|---|---|---|
| **Site** | | 27 | | |
| **Pas de « robots.txt »** | | 0 | | |
| | Google-Extended | 6 | 16,7 % | 18,5 % |
| | CCBot | 8 | 0,0 % | 29,6 % |
| | GPTBot | 13 | 0,0 % | 44,4 % |
| | ChatGPT-User | 12 | 0,0 % | 40,7 % |

Cette politique spécifique aux producteurs de contenus est cohérente avec leur modèle d'affaires. D'une part, les éditeurs vivent de leurs contenus. Les IA génératives permettent de venir concurrencer ces contenus, grâce à leurs modèles entraînés sur ces contenus, sans qu'aucun partage de revenus n'ait été négocié. La situation de dépendance est sensiblement différente de celle vécue, par exemple, entre Google et les éditeurs de presse (Ouakrat, 2020 ; Rebillard & Smyrnaios, 2010). En effet, il existe une dépendance mutuelle : Google a besoin des contenus des éditeurs pour alimenter ses bases de données, les éditeurs de presse ont besoin de Google pour que les utilisateurs retrouvent le contenu souhaité parmi les milliards de pages constituant le Web. Dans le cas des IA génératives, la relation de dépendance est à l'avantage des éditeurs dès lors que les producteurs d'IA génératives ont impérativement besoin de contenus produits par des humains pour maintenir la qualité de leur outil (Loukides, 2023). De même, un robot tel que ChatGPT-User permet de déléguer l'analyse des contenus à l'agent conversationnel. Il prive donc le site web d'une partie de son audience. Or cette dernière influence directement les revenus publicitaires des sites de presse (Schiff, 2006).

Cette politique de blocage contribue, en plus des actions en justice, à l'établissement d'un rapport de force favorable vis-à-vis des producteurs d'IA génératives. En effet, le blocage entraîne différentes conséquences dommageables à leur proposition de valeur. Premièrement, les blocages par protocole d'exclusion des

robots sont connues pour entraîner un biais dans l'information relayée par les moteurs de recherche (Sun, Zhuang, Councill & Giles, 2007) ; suffisamment suivies, ces politiques contribueront également aux biais dans les réponses fournies par les IA génératives. Deuxièmement, elles dégradent le service associé à la version payante de ChatGPT puisque les traitements automatisés depuis le client ChatGPT Plus (p. ex. synthèses de documents) deviennent impossibles une fois bloqué un des deux robots d'OpenAI. Des accords individuels commencent d'ailleurs à être contractés (p. ex. Axel Springer[8]), comme jadis entre la presse et les moteurs de recherche (Ouakrat, 2020).

Par contre, il est surprenant de constater que d'autres robots intervenant dans la production de jeux de données d'entraînement soient moins cités dans les fichiers « *robots.txt* ». C'est en particulier le cas de CCBot et de Google-Extended (ce constat est surtout valable pour les sites de presse française). Le CCBot permet la constitution du Common Crawl, notamment utilisé par de nombreux producteurs de LLM (voir par exemple les *datasets* des modèles hébergés sur Hugging Face), tandis que Google-Extended est utilisé par Google pour l'alimentation de Google Bard. Cela signifie que bloquer GPTBot mais pas CCBot n'empêchera pas l'alimentation du jeu de données d'entraînement utilisé par GPT. Plusieurs explications pourraient être proposées. Dans le cas de Google-Extended, la presse tend (difficilement) à normaliser ses relations avec Google, d'abord via des accords individuels, puis des accords sectoriels suite aux récentes évolutions légales dans l'Union européenne. De plus, Google utilise l'IA, d'une part, pour l'agent conversationnel Google Bard, d'autre part, pour son moteur de recherche (via *Search Generative Experience*), ce qui peut conduire les gestionnaires de sites à plus de prudence dans leur politique de blocage sélective. Concernant CCBot, le moindre blocage pourrait s'expliquer, d'une part, par une tolérance liée à l'usage du CCBot dans le monde de la recherche (incluant des projets sans lien direct avec les LLM commerciaux), d'autre part, par un manque de connaissance des jeux de données réellement utilisés par les producteurs d'IA génératives.

Ces résultats corroborent l'étude antérieure réalisée par Originality.ai (2023). Celle-ci relève cependant un taux de blocage sensiblement plus élevé pour le GPTBot (25.9% sur un Top 1000 mondial). Cette différence peut s'expliquer par la composition différente de l'échantillon de sites puisque les pratiques de blocage semblent varier fortement d'une catégorie de sites à une autre. Elle confirme par contre l'augmentation du blocage au fil des semaines ainsi qu'un blocage plus important de GPTBot comparativement à CCBot, à Google-Extended et à anthoropic-ai (le robot d'exploration associé au *chatbot* Claude). Ces blocages sélectifs sont susceptibles, comme pour les moteurs de recherche (Sun, 2008) d'induire des biais au niveau des contenus générés. Parmi les sites bloquant ChatGPT nous y trouvons notamment Amazon, Quora, NYTimes, Shutterstock, Wikihow et CNN. Nous pouvons y ajouter Stackoverflow[9]. Parmi les motivations possibles à ces blocages citons la protection de contenus originaux soumis au droit d'auteurs (journaux, stocks photos, QA), potentiellement liée à la défense de la qualité de partenariats antérieurs (p. ex. accord New York Times – Google ; Weatherbed, 2023), la lutte contre la contamination par des contenus hallucinés (QA) et la préservation d'avantages concurrentiels (p. ex. Amazon Bedrock).

[8] Voir https://openai.com/blog/axel-springer-partnership.

[9] Voir https://meta.stackoverflow.com/questions/421831/temporary-policy-generative-ai-e-g-chatgpt-is-banned.

## 5. Estimation des biais

L'approche utilisée dans l'article de Sun, Zhuang, Councill et Giles (2007) a été initialement conçue pour mesurer le biais des robots des moteurs de recherche. Cette méthode s'est avérée utile non seulement pour cet objectif initial mais aussi pour une analyse plus générale des fichiers « *robots.txt* » de différents sites web. Grâce à cette analyse, il est possible de quantifier le biais et de comparer la (dé)favorisation entre les différents moteurs de recherche mais aussi, dans notre cas, entre les différents robots d'exploration des IA génératives. La méthode se concentre sur deux aspects : le premier est le calcul du biais absolu d'un robot sur un site web spécifique, mesurant ainsi directement son comportement. Le second aspect évalue le biais relatif d'un robot par rapport à d'autres, en examinant les interactions sur un éventail de sites étudiés. Cette approche permet d'obtenir une vision à la fois directe et comparative du comportement des sites par rapport aux robots.

Deux types de biais sont calculés : le biais local et le biais global. Le biais local fournit une information absolue sur la (dé)favorisation d'un robot sur un site en particulier. Un robot est dit favorisé s'il peut accéder à un plus grand nombre de répertoires que le robot universel et inversement. Le robot universel étant définit comme tout robot qui ne correspond à aucun des noms *User-Agent* spécifiques dans le fichier « *robots.txt* ». L'algorithme permettant de calculer le biais d'un robot en particulier a donc besoin de trois informations : l'ensemble des répertoires du site, l'ensemble des répertoires accessibles par le robot universel et l'ensemble des répertoires accessibles par ce robot en particulier. L'ensemble des répertoires du site n'est pas une donnée accessible de façon évidente. Sun et ses co-auteurs approximent dès lors cet ensemble comme l'ensemble des répertoires présents dans un fichier « *robots.txt* ».

$$Biais(x) = \frac{N_{fav}(x) - N_{defav}(x)}{N} \quad (1)$$

La mesure du biais global, de façon absolue, donne une information utile. Cependant, elle reste difficile à interpréter et est limitée à la comparaison de robots sur un même site. La mesure du biais globale, basée sur cette dernière, permet de donner un indicateur relatif, normalisé entre -1 (biais défavorable) et +1 (biais favorable) pour un robot sur un ensemble de sites. Ce biais global se formule, pour un robot « *x* », comme indiqué dans l'équation (1), dans laquelle *N* représente le nombre de sites considérés, $N_{fav}(x)$, le nombre de sites pour lequel le biais est positif et $N_{defav}(x)$, le nombre de sites pour lequel le biais local est négatif. Le résultat représente donc la différence des proportions entre les sites qui favorisent le robot sélectionné et ceux qui le défavorisent. La valeur représente le pourcentage, en valeur absolue, de sites (de l'échantillon) qui favorisent (signe positif) ou défavorisent (signe négatif) le robot.

La mesure du biais global pose cependant un problème, en ce sens que la mise en œuvre du protocole d'exclusion des robots dans notre échantillon de site s'appuie très souvent sur une politique de liste blanche (blocage par défaut du robot universel puis autorisation sélective) ou de liste noire (autorisation par défaut du robot universel puis blocage sélectif). Or le blocage de la totalité d'un site passe par le blocage du répertoire racine (« / »), soit un seul répertoire. L'implémentation de la mesure du biais local par Sun et ses co-auteurs conduit donc à considérer comme favorisé un site dont la racine serait bloquée face au robot universel dont seuls

quelques répertoires seraient bloqués. Dans notre cas, lorsque les robots des IA génératives sont cités, nous avons par ailleurs vu que cela était pour les bloquer. Nous pouvons donc utiliser l'algorithme simplifié suivant pour calculer directement le biais global des robots d'exploration des IA génératives (cf. Algorithme 1).

*Algorithme 1. Calcul du biais global.*

```
Si robot pas cité :
  biais = 0
Si robot cité :
  Si robot universel bloqué :
    Si robot avec autorisation :
      biais = 1
  Sinon:
    Si robot cité avec blocage :
      biais = -1
    Sinon :
      biais = 0
```

La généralisation de cet algorithme peut être envisagée en considérant les combinaisons possibles de configuration des accès pour le robot universel et pour le robot spécifique considéré (cf. Tableau 4). En première approximation, nous avons considéré équivalents les accès partiels différents pour deux robots (cellules grisées). Ces valeurs (biais défavorable, neutralité, biais favorable) peuvent ensuite être combinées avec l'équation (1) afin d'obtenir une estimation du biais global pour chaque robot. Cette algorithme a été mis en œuvre avec la langage Python. Après un calcul manuel sur les sites de presse française, un calcul automatisé a été réalisé pour l'ensemble des sites (cf. Tableau 5). En pratique, les cas avec autorisation partielle représentent moins de 5 % des entrées analysées dans les fichiers « *robots.txt* ».

*Tableau 4. Détermination des biais locaux en fonction des modalités d'accès.*

| **Robot spécifique :** / **Robot universel :** | **Rien** | **Disallow (total)** | **Disallow (partiel)** | **Allow + Disallow** | **Allow (partiel)** | **Allow (total)** |
|---|---|---|---|---|---|---|
| **Rien** | 0 | -1 | -1 | -1 | -1 | 0 |
| **Disallow (total)** | 0 | 0 | 1 | 1 | 1 | 1 |
| **Disallow (partiel)** | 0 | -1 | 0 | 0 | 0 | 1 |
| **Allow + Disallow** | 0 | -1 | 0 | 0 | 0 | 1 |
| **Allow (partiel)** | 0 | -1 | 0 | 0 | 0 | 1 |
| **Allow (total)** | 0 | -1 | -1 | -1 | -1 | 0 |

Le calcul du biais global (cf. Tableau 5) concernant les robots d'exploration des IA génératives permet de mettre en évidence le traitement défavorable aux produits d'OpenAI (GPTBot et ChatGPT-User) comparativement à ceux de Google (Google Bard). Ce biais est même amplifié par le fait qu'OpenAI bloque les deux robots dès lors que l'un des deux robots est bloqué. Une fois cette règle intégrée, le biais pour GPTBot et ChatGPT-User s'élève en réalité à -0,41 (OpenAI) pour les sites de presse française. À titre de comparaison, les biais globaux pour Googlebot et Bingbot pour l'ensemble des sites s'élèvent respectivement à 0,07 et 0,05.

*Tableau 5. Calcul du biais global (exécution du 21-01-2024).*

| Robot | Biais global (presse française) | Biais global (tous les sites) |
|---|---|---|
| Google-Extended | -0,15 | 0 |
| CCBot | -0,15 | -0,14 |
| GPTBot | -0,37 | -0,23 |
| ChatGPT-User | -0,37 | -0,13 |

Cette différence de traitement est susceptible d'accroître les biais chez toutes ou partie de ces IA génératives. Si l'on s'appuie sur la typologie proposée par Ferrara (2023), plusieurs risques peuvent ainsi être identifiés. Premièrement, nous pouvons voir que certains contenus qualitatifs, tels que les articles de presse ou les articles scientifiques, étaient fréquemment bloqués. Si l'on admet que ces contenus, soumis à des règles déontologiques et à des processus de relecture, sont en moyenne moins sujets à des biais démographiques, culturels ou idéologiques, cela pourrait aboutir à des modèles davantage biaisés. Deuxièmement, au vu des différences de traitement entre agents, certaines IA pourraient se révéler davantage biaisées que d'autres. Nous constatons par ailleurs une corrélation inversée entre la popularité de l'outil associé au robot et le traitement favorable qui lui est réservé. Si Sun et al. (2007) constatent que les robots d'exploration des moteurs de recherche les plus populaires (en termes de parts de marché) sont les plus favorisés, l'inverse se produit avec les IA génératives, où l'outil le plus populaire (ChatGPT) fait l'objet du traitement le plus défavorable. Cela entraîne que l'outil le plus utilisé est aussi celui qui est le plus menacé par les biais liés aux déséquilibres dans les données d'entraînement. Troisièmement, les politiques de blocage ne sont pas nécessairement homogènes dans toutes les zones géographiques. Ces disparités pourraient conduire à des biais linguistiques, en particulier pour des langues déjà sous-représentées dans les jeux de données. Quatrièmement, le blocage plus important des robots d'OpenAI est partiellement compensé par l'accès plus large réservé à CCBot. OpenAI conserve dès lors un accès aux données via le Common Crawl. Par contre, cette différence handicape la mise à jour des données par OpenAI, ce qui peut introduire des biais temporels du fait des difficultés d'accès à certaines informations récentes.

*Tableau 6. Biais global par orientation politique et par robot.*

| Robot | Extrême gauche | Gauche | Centre gauche | Centre | Centre droit | Droite | Extrême droite | Complo-tiste |
|---|---|---|---|---|---|---|---|---|
| CCBot | 0,000 | -0,313 | -0,320 | -0,212 | -0,439 | -0,214 | 0,000 | 0,000 |
| GPTBot | -0,222 | -0,375 | -0,420 | -0,308 | -0,512 | -0,500 | 0,000 | -0.091 |

Pour développer ces éléments de discussion, nous avons généré avec ChatGPT (sous GPT-4) une base de données de sites de presse, pour chaque pays de l'Union européenne et le Royaume-Uni. Pour chaque site, nous disposons du nom, de l'URL, du pays, de la langue (codes ISO) et de l'orientation politique (suivant les catégories suivantes : extrême gauche, gauche, centre gauche, centre, centre droit, droite, extrême-droite). Notons que, par défaut, ChatGPT n'inclut jamais de site d'extrême-gauche ou d'extrême-droite dans ses réponses. Un *prompt* spécifique a donc été utilisé pour générer ces deux listes. ChatGPT refusant de créer une liste de journaux

d'extrême-droite puis générant finalement, après justification de l'usage de ces données (*sic*), une telle liste en changeant l'orientation (droite conservatrice), nous avons finalement créé une liste de site de journaux étiquetés à l'extrême-droite en recourant à l'agent conversationnel Le Chat proposé par Mistral. Les catégorisations ont été contrôlées par un échantillonage aléatoire. Au final, un ensemble de 190 fichiers robots (sur 206 sites listés) ont pu être explorés. Pour chaque site, et pour 6 robots (ccbot, gptbot, chatpt-user, google-extended, googlebot, bingbot), les biais locaux ont ensuite été calculés à l'aide du script Python (-1, 0 ou +1 ; cf. Tableau 4). Ces données calculées ont ensuite été traitées dans LibreOffice.org Calc pour estimer le biais global par pays, par langue et par orientation politique (cf. Tableau 6). Les valeurs de biais global permettent notamment de constater un moindre blocage du côté des journaux aux positionnements extrêmes (cf. Tableau 6). Les blocages varient également fortement par pays (biais global fortement défavorable, soit inférieur ou égal à -0,5, pour le Common Crawl concernant l'Allemagne, la Belgique, le Danemark, la Finlande, le Luxembourg et les Pays-Bas) et par langue (biais global fortement défavorable pour le néerlandais, le danois et le finnois). Cette exploration relative à des sources de données qualitatives (presse) permet donc de mettre en évidence les biais culturels, linguistiques et politiques induits par ces pratiques de blocage sélectif. Notons que le biais global est généralement davantage marqué pour GPT (gptbot) et nettement moins pour Bard/Gemini (google-extended). Un test supplémentaire sur 22 fichiers robots de sites complotistes proposés par Mistral a également révélé un biais global très faible (cf. Tableau 6).

La stimulation des biais par le recours différencié au protocole d'exclusion des robots pourrait par ailleurs conduire à un effet boule de neige. En effet, les LLM (*Large Language Model*) tels que GPT sont de plus en plus souvent utilisés comme générateurs de données pour l'entraînement, soit de SLM (*Small Language Model*) (Eldan & Li, 2023), soit de systèmes de *machine learning* plus classiques capables de réaliser, par exemple, des tâches de classification (Yu et al., 2023 ; Ye et al., 2022). Eldan et Li (2023) réussissent ainsi à développer un SLM (TinyStories) capable de rivaliser, sur le plan des capacités de génération de texte, avec un LLM (GPT-2) grâce à un effort sur la conception du jeu de données généré. La méthode s'appuie, d'une part, sur des *prompts* permettant de préserver les éléments essentiels du langage naturel dans le jeu de données généré, d'autre part, sur un *framework* basé sur GPT (GPT-4) permettant l'évaluation des textes générés par le SLM. Ye et ses co-auteurs (2022) montrent la possibilité de générer un jeu de données utilisable pour réaliser des tâches d'analyse de sentiments. Yu et ses co-auteurs (2023) font de même pour des tâches de classification. Ils montrent par ailleurs l'importance de la structure du *prompt* pour éviter de limiter la diversité des données générées et propager « *les biais systématiques hérités des LLM* ».

## 6. Conclusion

Cette recherche exploratoire a permis d'identifier les enjeux associés aux IA génératives de type *text-to-text* en matière de propriété intellectuelle puis d'analyser les stratégies de régulation mises en place par les gestionnaires de sites web. Leur conséquence a par ailleurs fait l'objet d'une discussion. Sept constats ont ainsi pu être dégagés.

Premièrement, et comparativement à des études antérieures telles que Sun (2008), l'utilisation du protocole d'exclusion des robots semble aujourd'hui

pratiquement généralisée. Deuxièmement, les gestionnaires de sites web se sont montrés réactifs face à l'émergence des IA génératives en intégrant des directives spécifiques aux robots d'exploration des IA génératives dans leurs fichiers d'exclusion de robots. Troisièmement, si l'on prend le Top 100 Alexa comme un étalon des pratiques courantes en matière d'exclusion des robots, l'activité des robots d'exploration des IA génératives apparaît peu ou prou problématique aux yeux des gestionnaires des sites web. Quatrièmement, les sites d'éditeurs de contenus (édition scientifique et, plus encore, presse) bloquent fréquemment les robots d'exploration des IA génératives, en particulier ceux clairement identifiés comme œuvrant à la création de vastes corpus d'entraînement. Cinquièmement, les stratégies de blocage de ces éditeurs se focalisent parfois trop sur OpenAI et négligent les outils de collecte associés à d'autres producteurs d'IA générative ou, plus grave, de jeux de données ouvertes utilisés par ces producteurs, ce qui réduit leur efficacité dans la constitution d'un rapport de force favorable vis-à-vis des producteurs d'IA génératives. Sixièmement, la réaction des gestionnaires de sites face aux IA génératives remet en lumière les limitations du protocole d'exclusion des robots et, d'une part, l'intérêt d'une désambiguïsation des consignes d'accès aux contenus en ligne (Yang & Liao, 2010 ; Sire, 2015), d'autre part, l'utilité de régulations spécifiques. Septièmement, les différences de politiques lors de l'exclusion de certains robots d'IA générative entraîne un risque accru de biais lors de l'entraînement de ces dernières.

Cette recherche ouvre la perspective d'une étude plus ambitieuse des biais potentiellement introduits par les pratiques de blocage des robots d'exploration des producteurs de LLM. Celle-ci passe par l'utilisation d'un annuaire de sites web, plus volumineux et diversifié, permettant d'analyser de manière systématique les biais induits, par exemple sur les plans culturels (p. ex. blocage différencié par pays ou par région) ou linguistiques (p. ex. blocage différencié par langue).

## 7. Références

Binctin, N. (2022). Droit de la propriété intellectuelle, droit d'auteur, brevet, droits voisin, marque, dessins et modèles. LGDJ. ISBN : 978-2275108339.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165. https://doi.org/10.48550/arXiv.2005.14165.

Casilli, A. (2019). En attendant les robots. Enquête sur le travail du clic. Seuil. ISBN : 978-2-0214-0188-2.

Chomsky, N., Roberts, I., & Watumull, J. (2023). Noam Chomsky: The False Promise of ChatGPT. The New York Times, 8. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

David, E. (2023). Now you can block OpenAI's web crawler. The Verge, 7 août 2023. https://www.theverge.com/2023/8/7/23823046/openai-data-scrape-block-ai.

Dignum, V. (2019). Responsible artificial intelligence: how to develop and use AI in a responsible way (Vol. 2156). Cham: Springer. ISBN : 978-3-030-30373-0.

Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., ... & Gardner, M. (2021). Documenting large webtext corpora: A case study on the colossal clean crawled corpus. arXiv preprint arXiv:2104.08758. https://doi.org/10.48550/arXiv.2104.08758.

Douet, M. (2023). Au Kenya, des « entraîneurs » de ChatGPT s'élèvent contre leurs conditions de travail. Le Monde, 19 octobre 2023. https://www.lemonde.fr/afrique/article/2023/10/19/au-kenya-des-entraineurs-de-chatgpt-s-elevent-contre-leurs-conditions-de-travail_6195464_3212.html.

Eldan, R., & Li, Y. (2023). TinyStories: How Small Can Language Models Be and Still Speak Coherent English?. arXiv preprint arXiv:2305.07759. https://doi.org/10.48550/arXiv.2305.07759.

Ferrara, E. (2023). Should chatgpt be biased? challenges and risks of bias in large language models. First Monday, Vol.28, N°11. https://doi.org/10.5210/fm.v28i11.13346.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30, 681-694. https://doi.org/10.1007/s11023-020-09548-1.

Galloway, S. (2018). The four - Le règne des quatre : la face cachée d'Amazon, Apple, Facebook et Google. Quanto. ISBN : 978-2889152469.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27. https://doi.org/10.48550/arXiv.1406.2661.

Hackett, C. (2023). OpenAI's ChatGPT plus: an electronic resources librarian's review. Journal of Electronic Resources Librarianship, 35(4), 299-304. https://doi.org/10.1080/1941126X.2023.2271373.

Heudin, J.-C. (2016). Comprendre le deep learning: Une introduction aux réseaux de neurones. Science-e-Book. ISBN : 979-1091245449.

Hu, K. (2023). ChatGPT sets record for fastest-growing user base - analyst note. Reuters, 2 février 2023. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

Ferrara, E. (2023). Should chatgpt be biased? Challenges and risks of bias in large language models. arXiv preprint arXiv:2304.03738. https://doi.org/10.5210/fm.v28i11.13346.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. D. L., ... & Sayed, W. E. (2023). Mistral 7B. arXiv preprint. https://doi.org/10.48550/arXiv.2310.06825.

Kalyan, K. S. (2023). A survey of GPT-3 family large language models including ChatGPT and GPT-4. Natural Language Processing Journal, 100048. https://doi.org/10.1016/j.nlp.2023.100048.

Layton, D. (2023). ChatGPT — Show me the Data Sources. Medium. https://medium.com/@dlaytonj2/chatgpt-show-me-the-data-sources-11e9433d57e8.

Loukides, M. (2023). Model Collapse: An Experiment. O'Reilly, 24 octobre 2023. https://www.oreilly.com/radar/model-collapse-an-experiment/.

Lucchi, N. (2023). ChatGPT: a case study on copyright challenges for generative artificial intelligence systems. European Journal of Risk Regulation, 1-23. https://doi.org/10.1017/err.2023.59.

Mattatia, F. (2017). Droit d'auteur & propriété intellectuelle dans le numérique. Eyrolles. ISBN : 978-2-416-00813-9.

McKenzie, S. & Arvanitis, L. (2023). Les "newsbots" montent au front : des sites d'actualité générés par l'IA se multiplient en ligne. NewsGuard, 01 mai 2023. https://www.newsguardtech.com/fr/special-reports/bots-ia-generative-sites/.

Originality.ai (2023). Websites That Have Blocked OpenAI's GPTBot CCBot Anthropic Google Extended - 1000 Website Study. https://originality.ai/ai-bot-blocking.

Ouakrat, A. (2020). Négocier la dépendance? Google, la presse et le droit voisin. Sur le journalisme, 9(1), 44-57. https://doi.org/10.25200/SLJ.v9.n1.2020.417.

Rebillard, F., & Smyrnaios, N. (2010). Les infomédiaires, au coeur de la filière de l'information en ligne: les cas de Google, Wikio et Paperblog. Réseaux, (2), 163-194. https://doi.org/10.3917/res.160.0163.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695). https://doi.org/10.48550/arXiv.2112.10752.

Schiff, F. (2003). Business models of news Web sites: A survey of empirical trends and expert opinion. First Monday. https://doi.org/10.5210/fm.v8i6.1061.

Singh, S. K., Kumar, S., & Mehra, P. S. (2023). Chat GPT & Google Bard AI: A Review. In 2023 International Conference on IoT, Communication and Automation Technology (ICICAT) (pp. 1-6). IEEE. https://doi.org/10.1109/ICICAT57735.2023.10263706.

Sire, G. (2015). Inclusion exclue: le code est un contrat léonin: Enquête sur la valeur technique et juridique du protocole robots. txt. Réseaux, (1), 187-214. https://doi.org/10.3917/res.189.0187.

Sun, Y. (2008). A comprehensive study of the regulation and behavior of web crawlers. Doctoral dissertation, The Pennsylvania State University. https://www.proquest.com/openview/22e8942f3aa45b2c043dba62f33ef3a1/1.

Sun, Y., Zhuang, Z., & Giles, C. L. (2007). A large-scale study of robots. txt. In Proceedings of the 16th international conference on World Wide Web (pp. 1123-1124). https://doi.org/10.1145/1242572.1242726.

Sun, Y., Zhuang, Z., Councill, I. G., & Giles, C. L. (2007). Determining bias to search engines from robots. txt. In IEEE/WIC/ACM International Conference on Web Intelligence (WI'07) (pp. 149-155). IEEE. https://doi.org/10.1109/WI.2007.98.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971. https://doi.org/10.48550/arXiv.2302.13971.

Tubaro, P., Casilli, A. A., & Coville, M. (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. Big Data & Society, 7(1), 2053951720919776. https://doi.org/10.1177/2053951720919776.

Weatherbed, J. (2023). The New York Times prohibits using its content to train AI models. The Verge, 14 avril 2023. https://www.theverge.com/2023/8/14/23831109/the-new-york-times-ai-web-scraping-rules-terms-of-service.

Yang, C., & Liao, H. J. (2010). Using the Robots. txt and Robots Meta tags to implement online copyright and a related amendment. Library hi tech, 28(1), 94-106. http://dx.doi.org/10.1108/07378831011026715.

Ye, J., Gao, J., Li, Q., Xu, H., Feng, J., Wu, Z., ... & Kong, L. (2022). Zerogen: Efficient zero-shot learning via dataset generation. arXiv preprint arXiv:2202.07922. https://doi.org/10.48550/arXiv.2202.07922.

Yu, Y., Zhuang, Y., Zhang, J., Meng, Y., Ratner, A., Krishna, R., ... & Zhang, C. (2023). Large language model as attributed training data generator: A tale of diversity and bias. arXiv preprint arXiv:2306.15895. https://doi.org/10.48550/arXiv.2306.15895.

Zirpoli, C. T. (2023). Generative Artificial Intelligence and Copyright Law. Congressional Research Service. https://crsreports.congress.gov/product/pdf/LSB/LSB10922.

# Cyberattaques : Impact des perceptions individuelles du risque dans l'activité de gestion de crise

## Une proposition d'analyse systémique, cognitive et ergonomique

**Marin François** [1,2], **Pierre-Emmanuel Arduin** [2], **Myriam Merad** [1]

1. *Université Paris-Dauphine, PSL, LAMSADE UMR CNRS 7243*
   *Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France*
   *marin.francois@dauphine.psl.eu*,myriam.merad@dauphine.psl.eu*
2. *Université Paris-Dauphine, PSL, DRM UMR CNRS 7088*
   *Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France*
   *pierre-emmanuel.arduin@dauphine.psl.eu*

RÉSUMÉ. *Dans cet article, nous proposons un méta-modèle des relations entre perceptions individuelles du risque et processus décisionnels collectifs dans le cadre de la gestion de crise cyber. Nous évaluons notre modèle à travers deux études de cas d'exercices de gestion de crise cyber au sein d'une organisation. Enfin, nous proposons un ensemble de cinq mesures complémentaires aux référentiels de gestion de crise cyber, permettant d'intégrer plus efficacement la prévention des effets d'entraînement liés à ces perceptions individuelles.*

ABSTRACT. *In this paper, we propose a meta-model of the relationships between individual risk perceptions and collective decision-making processes in the context of cyber crisis management. We evaluate our model through the case study of two cyber crisis management exercises within an organisation. Finally, we propose a set of five complementary measures for cyber crisis management frameworks to integrate more effectively the prevention of spillover effects linked to these individual perceptions during crisis management exercises.*

MOTS-CLÉS : *Gestion de crise, Cybersécurité, Risques, Gouvernance*

KEYWORDS: *Crisis Management, Cybersecurity, Risks, Governance*

## 1. Introduction

Prévoir les crises, que ce soit dans le domaine de la cybersécurité ou non, est un défi, si ce n'est une impossibilité (Ros, 2020, Renn, 2021), car les risques de crise sont *de-facto* complexes. La prévention des crises cyber se heurte ainsi à la difficulté de créer des modèles opérationnels didactiques (Boin, McConnell, 2007). Cependant, s'il n'est pas possible de prévoir le déroulement des évènements menant à une crise, il est néanmoins possible de formuler une suite de recommandations pour la neutralisation des effets d'entraînement liés à la prise de décision en situation de crise, *i.e.*, en évitant l'exacerbation de risques déjà présents (Fysarakis *et al*., 2022). Ainsi, le positionnement de l'équipe de gestion de crise est double : d'une part il lui revient de traiter techniquement les incidents symptomatiques de la crise, et d'autre part de veiller à ne pas empirer l'effet d'entraînement durant la prise de décision collective. À travers cette dualité, nous retrouvons le paradigme socio-technique propre aux systèmes d'information (Boaden, Lockett, 1991, Florent *et al*., 2019). Il faut donc distinguer les risques associés aux événements survenants en amont de la crise (« gestion de risques »), et les risques liés à l'activité de gestion de crise (« gestion de crise ») (Merad *et al*., 2011, Merad, Trump, 2020).

En matière de crise cyber, les référentiels généralistes tels que l'ISO-27001(-27002) et le NIST-CSF traitent principalement du contrôle des risques en amont, dans une approche interdisciplinaire et fonctionnelle. L'ISO-27001(-27002) ne mentionnent pas explicitement les bonnes pratiques pour contenir une crise, mais la mise en œuvre d'une méthode formelle d'analyse des risques pour la formalisation d'une procédure de gestion de crise. Dans le NIST Cybersecurity Framework (NIST-CSF), la gestion de crise est mentionnée sous les piliers « réponse » et « récupération » et se concentre principalement sur la gestion des incidents. À l'inverse, les référentiels spécialisés fournissent des informations détaillées sur les bonnes pratiques en matière de contrôle de la crise. L'ISO-22361, qui remplace la norme européenne CEN-TS 17091 pour la gestion de crise et la résilience depuis 2022 définit une crise selon trois caractéristiques: complexité, instabilité, incertitude. Les principes clés de la gestion de crise selon cette norme résident dans l'efficacité de la gouvernance, de la stratégie, de la gestion des risques, de la prise de décision et de la communication. Cette approche se retrouve dans la série des publications spécialisées NIST-SP-800, dont SP-800-30 (évaluation des risques), SP-800-61 (gestion des incidents) et SP-800-184 (gestion de la récupération). Le recours aux exercices de gestion de crise est une base commune de ces référentiels. Ces exercices permettent de tester les procédures opérationnelles et habituent les participants à la prise de décision collective en situation de crise, quand ces derniers subissent une forte pression psychologique individuelle (De la Garza, Weill-Fassina, 1995, Glendon, 1999, Merad *et al*., 2011).

En outre, les référentiels mentionnés (à l'exception de NIST SP-800-61) se limitent au périmètre de l'organisation, mais qu'en est-il de la gestion de crise au sein d'un réseau d'organisations ? (Boeke, 2018, Provan, Kenis, 2008). En effet, le système d'information et de connaissance (Arduin *et al*., 2015) globalisé est incompatible avec résilience « passive », centrée sur l'organisation, à travers une réponse dans l'urgence

à la dégradation des systèmes (Evrard Samuel, Ruel, 2013). Au contraire, la résilience « active », qui est un ajustement par l'apprentissage inter-organisationnel doit se faire par une gestion de crise centrée sur le réseau d'organisations. Nous n'avons cependant identifié aucun modèle des relations entre la perception individuelle des risques et l'activité de gestion de crise cyber, que ce soit au niveau de l'organisation ou au niveau du réseau. Nous proposons donc de répondre à la problématique suivante : quel est l'impact des perceptions individuelles du risque sur les processus de décision collective dans l'activité de gestion de crise cyber au sein d'un réseau d'organisations ?

Pour répondre à cette problématique, nous revenons d'abord sur les définitions de la gestion de crise et des risques systémiques (Hancock, 2002, Golandsky, 2016, Shrivastava, 1993), puis nous étudions les modèles de relations de perceptions individuelles et collectives et d'analyse ergonomique du risque proposés dans la littérature (Glendon, 1999, Glendon *et al*., 2016, Wilde, 1998, De la Garza, Weill-Fassina, 1995). Nous étudions par ailleurs l'impact de la gouvernance du réseau d'organisations sur la gestion des crises cyber (Boeke, 2018, Provan, Kenis, 2008), à travers les méthodes d'analyse micro-structurelle des réseaux de gouvernance (Provan, Kenis, 2008). Dans la troisième section de cet article, nous proposons un méta-modèle des relations entre perceptions individuelles du risque et processus de décision collective, mettant en évidence l'amplitude du facteur d'impact individuel $\alpha$ (Zuccaro *et al*., 2018) et le potentiel d'entraînement qu'il induit dans l'activité de gestion de crise cyber (Merad, Trump, 2020, Merad *et al*., 2011), à l'échelle du réseau d'organisation. Dans la quatrième section, nous évaluons le méta-modèle proposé à travers deux études de cas réalisées en 2022 et 2023. En nous appuyant sur les méthodes d'analyse ergonomique (De la Garza, Weill-Fassina, 1995), nous procédons à l'extraction de schémas types d'aggravation des risques. Dans la dernière partie de cet article, nous proposons cinq bonnes pratiques complémentaires aux référentiels de gestion de crise cyber pour mieux prendre en compte l'impact des perceptions individuelles sur l'aggravation des effets d'entraînement.

## 2. Revue de littérature

### *2.1. Crises Cyber*

Une crise cyber est un événement résultant d'incidents en cascade (Sherman *et al*., 2018) menaçant la sécurité de l'information et pouvant entraîner des dommages financiers, réputationnels et opérationnels graves (Hancock, 2002). Les référentiels de gestion de crise (pas nécessairement cyber) reposent sur des éléments techniques, organisationnels et scientifiques (Bénaben, 2016, Kulikova *et al*., 2012), mais se veulent généralistes et haut-niveau, adressant plus globalement la gestion de risque que la gestion de crise (Miller, Griffy-Brown, 2018). Dans l'ensemble, les concepts fondamentaux de la gestion de crise s'appliquent à la gestion des crises cyber, avec pour élément central la préparation (Hancock, 2002, Johansson, Härenstam, 2013, Kovoor-Misra *et al*., 2001).

La gestion des crises cyber implique d'abord l'anticipation des procédures – en portant attention particulière aux problèmes de sécurité théoriques (Mikolaj, 2005), puis pratiques, à travers la réponse et la récupération (Golandsky, 2016). La conformité réglementaire et la gestion de la réputation sont également à prendre en compte (Kulikova *et al*., 2012). La gestion efficace des crises cyber suppose une gestion efficace des connaissances et une communication adaptée (Johansson, Härenstam, 2013, Kulikova *et al*., 2012), la participation de multiples parties-prenantes (Lauras *et al*., 2015) pour une réponse rapide et coordonnée (Trimintzios *et al*., 2015).

L'évaluation et l'apprentissage sont cruciaux (Dawes *et al*., 2004) pour l'amélioration de la gestion de crise (Golandsky, 2016), au même titre que la construction de structures organisationnelles durables par les ressources internes, la formation des équipes et les investissements (Shrivastava, 1993). Les organisations peuvent apprendre des crises passées et aspirer au développement de leurs capacités de réponse (Bederna *et al*., 2017), pour passer d'une posture réactive à une posture anticipative (Shrivastava, 1993).

### *2.2. Nature des risques*

Les risques systémiques sont caractérisés par quatre composantes (Renn, 2011). *(1) Complexité*, résultant en des difficultés à établir des relations causales entre plusieurs événements et effets indésirables; *(2) Incertitude*, caractérisée par une forte variation statistique des observations, un environnement prône aux erreurs de mesure, par un manque de connaissance ou une forte indétermination (Renn, 2011, 2021, Renn *et al*., 2019, Van Asselt, 2000); *(3) Ambiguïté*, sous-jacente à la variabilité dans l'interprétation logique des événements observés et l'incertitude (Renn *et al*., 2011), et *(4) Effet d'entraînement* au-delà de l'environnement des sources de risques (Kasperson *et al*., 2003).

Selon cette définition, les risques de crise cyber sont *de facto* des risques systémiques (Davis, 2005, Forscey *et al*., 2022, Sommer, Brown, 2011). Schweizer (2021), German Advisory (2018) proposent de qualifier les risques systémiques selon un modèle d'aide à la décision multicritère prennant en compte les facteurs d'évaluation suivants : étendue et persistence des dommages, ubiquité, réversibilité, latence, violation de l'équité et potentiel de mobilisation.

Ainsi, c'est parce qu'ils possèdent ces attributs très spécifiques qu'ils diffèrent des risques « classiques », connus sous le nom de risques « idiosyncratiques ». Ces derniers sont potentiellement quantifiables et prévisibles, ce qui n'est pas le cas pour les risques systémiques. Le tableau 1 résume les différences entre ces deux types de risques.

### *2.3. Perceptions du risque*

L'un des modèles consensuels d'analyse des perceptions individuelles du risque est proposé par Glendon (1999), Hale, Glendon (1987). Cependant, dans sa première

TABLEAU 1. *Risques Systémiques vs. Idiosyncratiques*

| Risques Systémiques | Risques Idiosyncratiques |
|---|---|
| Grande Complexité | Effet Causal Direct |
| Réponse Non Linéaire | Réponse Linéaire |
| Forte Stochasticité | Stochasticité Limitée |
| Distribution loi de puissance | Distribution Normale |
| Indétermination | Déterminé |
| Effet d'entraînement extrême | Faible effet d'entraînement |

version, le modèle ne prend pas en considération la perception individuelle du risque lorsque l'individu fait partie d'un groupe. Le deuxième modèle d'interprétation des risques ultérieurement proposé par Glendon *et al*. (2016) introduit des risques spéculatifs (relatifs aux domaines de l'égo et du social), conciliant la Théorie de l'Homéostasie des Risques (RHT) de Wilde (1998) avec la théorie de la prise de décision organisationnelle – replaçant l'individu au sein du groupe (voir Tableau 2).

Ce modèle implique les processus de « *hot-cognition* » et « *cold-cognition* » à travers quatre questions sources impactant la psychologie de l'individu. Le concept de « *hot-cognition* » et « *cold-cognition* », détaillé dans Glendon (1999), revient notamment à jauger l'impact des émotions sur le processus cognitif. De Smidt, Botzen (2018) ont étudié, pour le cas précis de l'évaluation de risques cyber, comment les biais de jugement des individus occupant différents rôles de gestion de risque cyber impactaient leurs perceptions, montrant que les décideurs techniques surestiment les probabilités d'attaque et sous-estiment les impacts financiers. En outre, la vulnérabilité perçue, les croyances en compétences propres (Debb, McClellan, 2021), les aptitudes personnelles (Kostyuk, Wayne, 2021), interpersonnelles et l'heuristique d'affect (Skagerlund *et al*., 2020, Van Schaik *et al*., 2020) influencent également la perception du risque et donc la gestion de crise.

Pour extraire des schémas types de situation d'aggravation à travers l'interprétation collective du risque, De la Garza, Weill-Fassina (1995) s'appuient sur les méthodes d'analyse systémique et cognitive. Historiquement, la première approche se concentre sur les caractéristiques environnementales du risque, tandis que la seconde considère les schémas d'interprétation de l'individu. Les auteurs s'appuient sur la première version du schéma d'interprétation et de réaction aux risques (Glendon, 1999, Hale, Glendon, 1987), combiné à l'approche systémique. Dans cet article, nous nous appuyons sur le « cadre combiné d'analyse ergonomique » ainsi proposé par les auteurs, cependant, nous utiliserons le modèle de Glendon *et al*. (2016), puisque celui-ci prend en compte la dimension collective.

### *2.4. Réseaux de Gouvernance*

Pour Provan, Kenis (2008), les réseaux sont des ensembles regroupant trois organisations indépendantes ou plus avec des objectifs individuels et collectifs, présentant des dynamiques spontanées, mandatées ou contractuelles. Les auteurs proposent des

TABLEAU 2. *Perception individuelle vs. collective du risque*

| Individuelle | Collective |
|---|---|
| « *Cold cognition* » | « *Cold & Hot cognition* » |
| Risques purs | Risques purs & spéculatifs |
| Adaptations choisies | Culture, règles, politiques |
| Boucle unique fermée | Boucles multiples |

éléments pour l'analyse des modes de gouvernance dans les réseaux d'organisations, où deux approches prévalent : l'approche micro-structurelle et l'approche du réseau en tant que forme de gouvernance (Provan, Kenis, 2008). Dans la permière, le réseau est analysé à travers les relations et les caractéristiques des individus qui le composent, dans la seconde, le réseau est analysé « à l'échelle », ce qui signifie que l'unité de référence est celle du réseau, et non plus des organisations qui le composent.

Provan, Kenis (2008) proposent un cadre tenant compte de différentes configurations micro-structurelles et incluant également la dynamique du réseau à l'échelle. Trois modes distincts de gouvernance des réseaux sont indentifiés : (1) Réseau gouverné par les participants (« *Shared Governance* »), (2) Réseau gouverné par une organisation meneuse (« *Lead Agency* ») et (3) Réseau gouverné par une organisation administrative annexe (« NAO »). Pour Provan, Kenis (2008) toujours, l'adaptation entre le mode de gouvernance et les caractéristiques micro-structurelles du réseau définit la performance ou l'échec de la gouvernance. Les organisations doivent choisir un mode de gouvernance adapté aux caractéristiques du réseau : niveau de confiance au sein du réseau, nombre de participants, force du consensus sur les objectifs, besoin en compétences techniques, *etc.*... Puis, arbitrer parmi plusieurs avantages et inconvénients liés au mode de gouvernance choisi, comme l'inclusivité de la prise de décison et la réactivité, la légitimité interne des choix stratégiques et la légitimité externe, la stabilité des processus et la flexibilité.

Tous ces éléments influencent directement la résilience du réseau et des organisations qui le composent, comme le montre Boeke (2018), en appliquant la méthode d'analyse de Provan, Kenis (2008) pour étudier les différences structurelles dans les modes de gouvernance des états au sein du consortium européen de collaboration pour la cybersécurité (EU-CyCLONe).

## 3. Méta-modèle et méthode d'analyse

Maintenant que nous avons une vue d'ensemble des relations liant la perception individuelle du risque, la perception collective et les conditions de performance de la gouvernance en réseau, nous sommes en mesure de proposer un méta-modèle reliant ces éléments. Le modèle proposé est présenté dans la Figure 1.

**(A)** Dans notre méta-modèle, les biais de jugement, par exemple liés à l'expertise, ont un impact direct sur les perceptions individuelles du risque (Skagerlund *et al.*,

2020, Van Schaik *et al*., 2020), au même titre que l'absence de connaissances (Arduin *et al*., 2015).

**(B)** Nous stipulons ensuite que les perceptions individuelles du risque, à travers les « *hot-cognition* » et « *cold-cognition* », (Glendon, 1999, Glendon *et al*., 2016, Hale, Glendon, 1987) modifient l'appétence au risque et l'interprétation des signaux en situation de crise (Wilde, 1998). L'évaluation du risque au niveau collectif, tenant compte des facteurs cognitifs spécifiques aux perceptions individuelles au sein du groupe (Glendon, 1999) est donc différente de celle des individus.

**(C)** Lorsque ces organisations forment des réseaux, ces derniers se caractérisent par l'alignement entre le mode de gouvernance choisi et les caractéristiques structurelles du réseau (Provan, Kenis, 2008). Selon Provan, Kenis (2008), cet alignement agit sur la perception du risque par les individus. Par exemple, un niveau de méfiance élevé au profit d'une gouvernance partagée aura un impact direct sur l'appétence au risque des individus, ce qui pourra se caractériser à terme, par l'adoption de règles, référentiels et normes communes et une incitation au contrôle entre les parties prenantes. C'est également cet alignement entre structure et mode de gouvernance qui impactera la performance du réseau en situation de gestion de crise, en impactant par exemple sa réactivité.

**(D)** À leur tour, les performances du réseau auront un impact sur les individus. *Ex-post*, les individus considéreront la structure de gouvernance du réseau à la lumière des décisions prises et de leur interprétation du déroulement des événements, les amenant à réévaluer leurs perceptions. Cela passe notamment par l'interprétation retrospective des décisions prises par les individus et le réseau par rapport aux conséquences de la crise et de sa gestion.

**(E)** Par ailleurs, les individus percevront les événements affectant d'autres participants de leur environnement (par exemple, une crise similaire chez un concurrent) et ajusteront leurs perceptions individuelles en tenant compte de ces éléments, ce sont les risques subjectifs.

**(F)** Enfin, nous stipulons que naturellement, le mode de gouvernance évoluera par la transformation du réseau vers un alignement optimal (Provan, Kenis, 2008) et que tous ces éléments sont soumis à l'incitation au changement fournit par la technologie (Provan, Kenis, 2008).

Notre méta-modèle permet donc de combiner les modèles existants pour un cadre d'analyse combiné incluant les perceptions individuelles, collectives, et entre plusieurs organisations dans le cadre d'une activité de gestion de crise. Par ailleurs, il permet d'analyser le caractère évolutif des pratiques de gestion de crise, en prenant en compte l'apprentissage organisationnel. Cette proposition n'existe – à notre connaissance – pas dans la littérature. Sans cette proposition, nous devrions analyser les données collectées au regard de trois grilles différentes, puis conjuger les résultats. Nous devrions par ailleurs faire ce travail pour les deux années, sans garantie de pouvoir lier les observations d'une année avec la suivante. Ainsi, en conjugant les éléments issus de la littérature dans un cadre combiné d'analyse, nous pouvons désormais utiliser les fac-

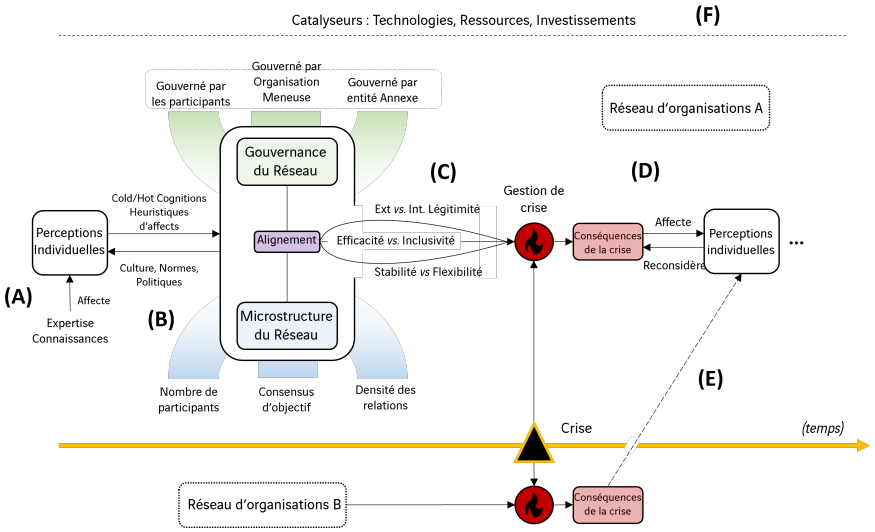



FIGURE 1. *Meta-modèle des relations entre le mode de gouvernance, les perceptions individuelles et l'activité de gestion de crise au sein du réseau d'organisations*

teurs listés et leurs interactions afin de valider ou non leur influence sur l'activité de gestion de crise et son évolution d'une année sur l'autre.

Nous allons maintenant évaluer ce méta-modèle (voir Figure 1) au regard de l'ensemble des données recueillies sur le terrain. Nous utilisons deux enregistrements d'une unité opérationnelle de gestion de crise cyber au sein d'un réseau d'organisations du secteur de l'énergie. Le premier enregistrement date du troisième trimestre de 2022, le second d'un an plus tard. Les protagonistes sont experts dans différents domaines (terminaux, réseaux, sécurité, applications, ERP, ...). Chaque enregistrement dure environ une demi-journée. Le tableau 3 résume les caractéristiques des enregistrements.

TABLEAU 3. *Caractéristiques des enregistrements*

| Caractéristique | 2022 | 2023 |
|---|---|---|
| Scénario | Rançongiciel | Compromission prestataire |
| Profils | Experts Techniques | Experts Techniques |
| Chronologie | Q3 2022 | Q3 2023 |
| Lieu | Salle de crise | Salle de crise + À distance |
| Durée | Demi-journée | Demi-journée |
| # de participants | 25 | 79 |
| % de participants formés à la gestion de crise | 100% | environ 30% |
| Nombre d'Organisations | 1 | 45 |
| Structure | Organisation Unique | Réseau d'Organisations |

### *3.1. Méthode d'analyse ergonomique*

Nous analysons les pratiques de gestion de crise telles qu'elles sont définies dans les référentiels par rapport aux facteurs de réussite ou d'échec tels que définis dans notre méta-modèle. Une représentation synthétique de ce mode d'analyse, tel que repris de De la Garza, Weill-Fassina (1995) est présentée dans la Figure 2.

Dans les lignes d'un tableau nous positionnons les éléments définis dans les référentiels spécialisés, et les colonnes suivantes sont divisées en groupes de séquences (T1, T2, ...) correspondant aux phases de l'exercice. Les séquences identifiées ont été tirées des travaux de préparation des exercices de crise, de manière à *(i)* s'assurer que les *stimuli* sont cohérents (communications envoyées, évènements majeurs décrits), et *(ii)* que les réponses possibles à ces *stimuli* constituent des apprentissages souhaitables. Une colonne est définie pour chaque facteur de performance/risque tels qu'identifiés dans notre méta-modèle. Pour chaque séquence d'actions (par exemple, E1/T1 (Exercice 1, Séquence 1) - Incident de fraude au paiement signalé), nous indiquons *(a)* si la tâche de gestion de crise est effectuée telle que définit dans les référentiels, et *(b)* si l'un des facteurs de performance/risque a impacté la tâche.

Les perceptions individuelles sont ainsi analysées au travers du discours des participants en lien avec les facteurs définis en Tableau 2, dans une approche descriptive. Par exemple, durant la première séquence clé de l'exercice 2022 (E1-T1), la tâche « NIST CSF - DT / ISO22361: Les évenements sont analysés pour anticiper les actions adverses » est effectuée, mais les échanges qui s'en suivent débouchent sur deux analyses divergentes par deux parties de l'équipe de gestion de crise. L'analyse choisie est alors définit par l'opérateur avec le plus haut niveau hierarchique. La compréhension des activités adverses est donc notée dans notre tableau comme « influencée par le facteur *structure organisationnelle et hiérarchique* ».

Cette méthode d'analyse est calquée sur la méthode d'analyse ergonomique proposée par De la Garza, Weill-Fassina (1995), à la différence que nous n'utilisons pas les mêmes représentations opérationnelles (notre méta-modèle prend en compte les risques subjectifs) ni les mêmes unités pour les séquences de temps (dans l'article original, les heures sont utilisées, nous utilisons les séquences).

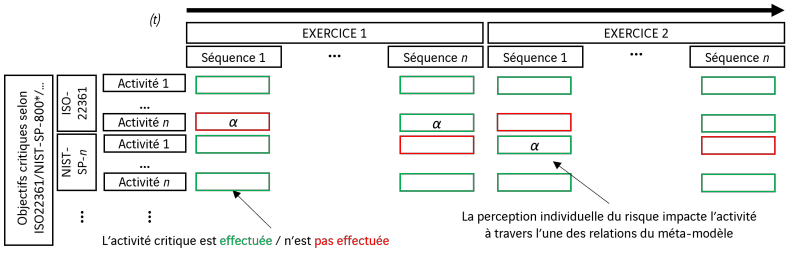



FIGURE 2. *Structure utilisée pour l'analyse ergnonomique, inspirée de De la Garza, Weill-Fassina (1995)*

## 4. Étude de cas

En analysant les données collectées, nous avons pu comparer à travers l'analyse descriptive du discours des participants et de leurs comportements, d'une année sur l'autre et comparer les mesures prises avec celles suggérées dans les référentiels. Nous avons pu valider la présence de facteurs influents tels que définis dans notre méta-modèle, notamment en identifiant des signes très concrets liés au processus de « *hot cognition* », à l'impact significatif du mode de gouvernance de la structure sur la gestion de crise, à l'importance de la confiance et de la prise de décision partagée dans les processus décisionnels sous pression, à l'impact des heuristiques d'affect et aux limites de l'efficacité des exercices de crise sur les participants. Les séquences analysées sont présentées dans le Tableau 4.

### *4.1. Impacts des perceptions individuelles du risque*

La structure de l'équipe opérationnelle, avec un leadership fort guidant les autres équipes, correspond à un mode de gouvernance par Organisation Meneuse (« *Lead Agency* », Provan, Kenis (2008)). La prise de décision est en conséquence centrale, extrêmement rapide dans cette organisation. Cependant, cela a également conduit à ce que les suggestions d'autres membres, qui se seraient avérées efficaces, ne soient pas prises en compte. Au cours du premier exercice, l'équipe opérationnelle centrale a perdu beaucoup de temps en ne tenant pas compte d'une suggestion d'investigation formulée par un protagoniste. Il est ensuite apparu que cette recommandation aurait permit de contenir rapidement la crise. Bien qu'au début de l'exercice, la prise de décision soit soumise à la validation partagée, à mesure que l'exercice progresse, celle-ci dérive vers une prise de décision directe et centrale. Il est fait mention explicite que les décideurs seraient tenus responsables de certaines décisions si elles étaient remises en question en dehors de l'organisation, assurant ainsi la légitimité externe. Presque tous les techniciens impliqués ont fait preuve d'heuristiques d'affect lors de l'analyse des impacts potentiels et des délais de réponse associés aux systèmes dont ils ont la responsabilité. Nous avons par exemple constaté que les estimations de temps de chargement étaient jusqu'à quatre fois plus élevées que les temps avérés. Enfin, nous avons identifié des limites à la pratique des exercices de gestion de crise. Par exemple, lorsque plusieurs protagonistes ont identifié des soucis de cohérence dans les séquences, cela les a incité à réduire leur implication dans le jeu. Nous avons également noté que l'absence d'impact réel des décisions des joueurs sur le scénario les a progressivement démobilisé et les a amené à devenir de moins en moins impliqués. L'évolution globale de ce processus est présentée dans la Figure 3.

### *4.2. Apprentissage au sein du réseau d'organisation*

Au cours du deuxième exercice, nous avons observé une évolution de la prise en compte de l'impact des décisions dans la gestion de crise. Nous avons pour cela comparé les processus décisionnels par rapport à l'année précédente au sein de l'orga-

TABLEAU 4. *Séquences des exercices*

| Phase | Événements principaux |
|---|---|
| EX 1 - T1. Découverte de l'intrusion | – Activités suspectes provenant du serveur SRV1 vers SRV2<br>– Renseignements révèlent une campagne active d'exploitation 0-day sur SRV1<br>– Les enquêtes sur SRV1 révèlent une exploitation 0-day |
| EX 1 - T2. Rançongiciel | – Les journaux SRV2 indiquent des tentatives de connexion infructueuses d'un administrateur<br>– Le SRV1 compromis est isolé<br>– Le compte utilisé pour les tentatives SRV1 vers SRV2 est bloqué<br>– Des problèmes de connexion sont apparus pour un petit groupe d'utilisateurs ERP<br>– L'investigation révèle que les comptes associés n'ont plus suffisamment de droits pour se connecter à ERP<br>– D'importantes modifications de l'annuaire ont été détectées sur SRV2 et le compte administrateur a été modifié<br>– Plusieurs utilisateurs signalent des échecs de connexion ERP<br>– Le département financier alerte sur des autorisations de paiement suspectes pour des montants élevés<br>– Un rançongiciel a été détecté sur plusieurs terminaux<br>– Une capture d'écran de la discussion de l'équipe de crise a été publiée sur Twitter avec le message « un coup d'avance »<br>– Plusieurs applications sont chiffrées et inutilisables<br>– Le compte utilisé pour propager le rançongiciel a été identifié<br>– Un e-mail contenant une demande de rançon a été reçu |
| EX 1 - T3. Planification de la récupération | – Les systèmes ERP ont été arrêtés<br>– Tout accès externe au système d'information a été fermé<br>– La diffusion du rançongiciel est terminée, 50 % des terminaux ont été chiffrés<br>– Une partie de l'ERP est chiffrée<br>– Plusieurs fournisseurs contactent les services financiers pour se plaindre de retards de paiement |
| EX 1 - T4. Récupération | – Les services de SRV1,SRV2 et de réseau sont reconstruits<br>– Antivirus est redéployé sur l'ensemble des terminaux<br>– Applications et bases de données sont restaurées à l'aide de sauvegardes<br>– Tous les ordinateurs impactés doivent être remis à neuf<br>– La communication avec les clients et fournisseurs est effectuée |
| EX 2 - T1. Accès initial | – Plusieurs utilisateurs de la direction ont reçu un courriel frauduleux demandant le paiement de 800 000£<br>– Alertes multiples sur le poste de travail de l'utilisateur émetteur du courriel, le compte est compromis<br>– Plusieurs entrées de données ERP ont été modifiées<br>– Un deuxième compte compromis a été confirmé sur le poste de travail identifié |
| EX 2 - T2. Modification des données métier | – Modification confirmée des données pour au moins deux zones géographiques<br>– L'investigation révèle que la compromission date d'une semaine<br>– Confirmation que l'e-mail de frauduleux a été envoyé par un fournisseur externe<br>– Alertes multiples sur plusieurs terminaux<br>– Confirmation du gel de modifications des serveurs |
| EX 2 - T3. Mode de blocage | – Modification confirmée des données ERP et irréversibilité constatée<br>– La direction financière demande le blocage des paiements<br>– Rançongiciel détecté par l'EDR sur plusieurs terminaux<br>– Plusieurs responsables financiers ont reçu le même courriel de demande de rançon<br>– Les alertes de rançongiciel sont confirmées faux positifs |
| EX 2 - T4. Rançon demandée | – E-mail demandant une rançon reçu, fichier de 145 Go supposément récupéré et menacé d'être publié<br>– Plusieurs protagonistes ont été contactés par des journalistes |

nisation meneuse. Les protagonistes sont à deux joueurs près les mêmes que l'année précédente, avec les mêmes responsabilités respectives, ce qui facilite notre comparaison. La différence principale dans la structure du réseau est l'ajout de nouvelles organisations « suiveuses » de l'organisation meneuse présente depuis le premier exercice. Dans la mesure où les joueurs savent que les premiers signaux faibles observés conduiront à une situation de crise, ils ont résisté à l'incitation à activer directement le protocole de crise et se sont contentés de traiter l'incident source. Mention explicite est également faite des exercices passés, et en particulier de la nécessité de tenir un registre rigoureusement mis à jour des actions entreprises et d'alerter si des informations sont perdues lors de la prise d'une décision. Le responsable de l'organisation meneuse a également explicitement demandé aux joueurs de partager leurs commentaires et intuitions avec lui tout au long de l'exercice, mettant en exergue une volonté d'incorporer une prise de décision partagée. De manière générale, nous avons identifié deux domaines principaux de changement entre les deux exercices annuels : une communication améliorée et, surtout, une prise de décision partagée. Il semble que l'équipe opérationnelle ait réussi à travailler sur un mode de gouvernance qui lui permet de maintenir un certain degré d'efficacité tout en incluant de manière plus efficace les intuitions et l'expertise des autres membres. Nous avons par ailleurs noté une augmentation du biais de confiance au cours de cet exercice, avec une remise en question extrêmement importante de la véracité des informations, généralement suivie d'opérations de vérification. Cependant, cette approche rigoureuse de la vérification des informations a conduit à plusieurs reprises à un gaspillage inutile de temps. Aucune stratégie optimale pour ce point n'est mentionnée dans les directives de gestion de crise.

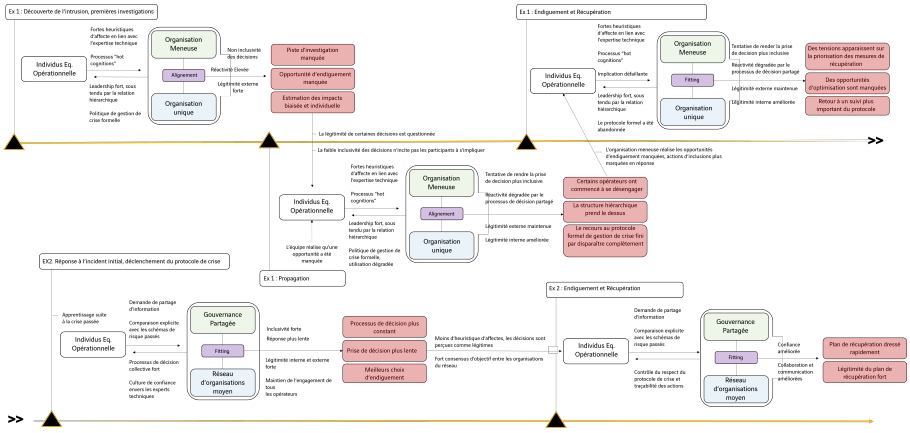



FIGURE 3. *Évolution de la gouvernance en lien avec les perceptions individuelles du risque*

## 5. Conclusions et perspectives

Nous avons proposé dans cet article un méta-modèle illustrant les relations complexes entre la perception individuelle du risque et la prise de décision collective dans le cadre de la gestion de crise cyber. Par deux études de cas, nous avons confirmé l'applicabilité de notre méta-modèle et démontré son efficacité pour identifier des schémas typiques d'aggravation du risque. L'impact des perceptions individuelles du risque émerge comme un élément significatif influençant l'efficacité des activités de gestion de crise, mais les normes actuelles de gestion de crise ne parviennent souvent pas à reconnaître et à aborder pleinement ce facteur. En intégrant explicitement l'interaction entre les perceptions individuelles du risque, la prise de décision collective et l'efficacité des modes de gouvernance du réseau lors de la gestion des crises cyber, ce méta-modèle offre une perspective nuancée qui peut être utilisée pour analyser la pratique de la gestion de crise au sein d'un réseau d'organisations. En réponse à la problématique de cet article, nous avons formulé ci-dessous cinq propositions didactiques pour la gestion de crise, au regard de notre analyse.

**Leadership** Si un leadership fort est crucial, il faut veiller à ce qu'il n'éclipse pas les suggestions des parties prenantes. Il faut ainsi encourager une communication ouverte et créer un environnement où tous les membres de l'équipe se sentent habilités à exprimer leurs perspectives.

**Heuristiques d'affects** Lors de la réalisation d'analyses d'impact, il convient de prendre en compte les heuristiques d'affects qui peuvent déformer les perceptions des analystes, en lien avec leurs spécialisations et les risques subjectifs perçus. Pour cela, il est possible de recourir à un outillage d'analyse d'impact partagé, connu et accepté des participants, faisant autorité en la matière et défini en amont de la crise.

**Mode de gouvernance** Adopter un modèle « Organisation Meneuse » pour une gouvernance efficace du réseau permet une réponse rapide aux incidents, l'idéal est de mettre en place des accords de décision partagée ou des points de validation réguliers. Cette inclusion permet non seulement une contribution collaborative, mais renforce également la légitimité interne et la confiance des participants sans compromettre l'efficacité globale.

**Culture et expérience** Il est pertinent de mentionner explicitement les crises passées lors des discussions pour aider à la visualisation et à la contextualisation. La référence à des instances spécifiques peut fournir un cadre tangible pour comprendre les défis potentiels et les solutions, en fournissant un référentiel partagé des schémas types de risques, améliorant la capacité de l'équipe à naviguer efficacement dans la crise.

**Confiance et légitimité** Les responsables d'équipes meneuses doivent donner la priorité à la légitimité interne et externe des décisions. En interne, il faut s'assurer que les décisions correspondent aux valeurs et aux objectifs de l'équipe, favorisant la confiance et la légitimité. En externe, il faut communiquer les décisions de manière transparente pour renforcer la légitimité des actions entreprises.

Dans nos futurs travaux, nous allons collecter les données d'un troisième exercice de crise afin de valider une nouvelle fois le méta-modèle proposé, en tentant d'anticiper les transformations du réseau que celui-ci pourrait induire. Par ailleurs, nous souhaiterions implémenter les recommandations formulées dans le plan de formation des opérateurs de l'organisation meneuse afin de vérifier leur efficacité. Aussi, nous souhaitons mettre en place un méthode d'analyse qualitative des perceptions individuelles plus complète que l'analyse descriptive proposée ici, notamment par la mise en place d'entretiens avec les participants. Enfin, nous souhaitons définir des marqueurs précis de l'influence technologique en tant que catalyseur des dynamiques des prise de décision, afin de vérifier que ces derniers ont un impact significatif, tels que nous l'avons stipulé dans notre méta-modèle.

## Bibliographie

Arduin P.-E., Grundstein M., Rosenthal-Sabroux C. (2015). *Système d'information et de connaissance* (vol. 4). ISTE Group.

Bederna Z., Rajnai Z., Szadeczky T. (2017). Further strategy analysis of cybersecurity incidents. *Land Forces Academy Review*, vol. 26, n° 3, p. 251–260.

Bénaben F. (2016). A formal framework for crisis management describing information flows and functional structure. *Procedia Engineering*, vol. 159, p. 353–356.

Boaden R., Lockett G. (1991). Information technology, information systems and information management: definition and development. *European Journal of Information Systems*, vol. 1, n° 1, p. 23–32.

Boeke S. (2018). National cyber crisis management: Different european approaches. *Governance*, vol. 31, n° 3, p. 449–464.

Boin A., McConnell A. (2007). Preparing for critical infrastructure breakdowns: the limits of crisis management and the need for resilience. *Journal of contingencies and crisis management*, vol. 15, n° 1, p. 50–59.

Davis B. J. (2005). Prepare: seeking systemic solutions for technological crisis management. *Knowledge and Process Management*, vol. 12, n° 2, p. 123–131.

Dawes S. S., Cresswell A. M., Cahan B. B. (2004). Learning from crisis: Lessons in human and information infrastructure from the world trade center response. *Social Science Computer Review*, vol. 22, n° 1, p. 52–66.

De la Garza C., Weill-Fassina A. (1995). Méthode d'analyse des difficultés de gestion du risque dans une activité collective: l'entretien des voies ferrées". *Safety Science*, vol. 18, n° 3, p. 157–180.

Debb S. M., McClellan M. K. (2021). Perceived vulnerability as a determinant of increased risk for cybersecurity risk behavior. *Cyberpsychology, Behavior, and Social Networking*, vol. 24, n° 9, p. 605–611.

De Smidt G., Botzen W. (2018). Perceptions of corporate cyber risks and insurance decision-making. *The Geneva Papers on Risk and Insurance-Issues and Practice*, vol. 43, p. 239–274.

Evrard Samuel K., Ruel S. (2013). Systèmes d'information et résilience des chaînes logistiques globales. *Systèmes d'information et management*, vol. 18, n° 1, p. 57–85.

Florent B., Nicolas M., Marchand A. L., Colin B. (2019). Cyber attaques: Organiser la confiance. In *Epic*.

Forscey D., Bateman J., Beecroft N., Woods B. (2022). *Systemic cyber risk: A primer*. Carnegie Endowment for International Peace.

Fysarakis K., Mavroeidis V., Athanatos M., Spanoudakis G., Ioannidis S. (2022). A blueprint for collaborative cybersecurity operations centres with capacity for shared situational awareness, coordinated response, and joint preparedness. In *2022 ieee international conference on big data (big data)*, p. 2601–2609.

German Advisory C. for. (2018). Strategies for managing global environmental risks.

Glendon I. (1999). Management of risks by individuals and organisations. *Safety Science Monitor*, vol. 3, n° 4, p. 2–11.

Glendon I., Clarke S., McKenna E. (2016). *Human safety and risk management*. Crc Press.

Golandsky Y. (2016). Cyber crisis management, survival or extinction? In *2016 international conference on cyber situational awareness, data analytics and assessment (cybersa)*, p. 1–4.

Hale A. R., Glendon I. (1987). *Individual behaviour in the control of danger*. Elsevier Science.

Hancock B. (2002). Security crisis management—the basics. *Computers & Security*, vol. 21, n° 5, p. 397–401.

Johansson A., Härenstam M. (2013). Knowledge communication: a key to successful crisis management. *Biosecurity and bioterrorism: biodefense strategy, practice, and science*, vol. 11, n° S1, p. S260–S263.

Kasperson R. E., Pidgeon N. F., Slovic P. (2003). *The social amplification of risk*. Cambridge University Press.

Kostyuk N., Wayne C. (2021). The microfoundations of state cybersecurity: Cyber risk perceptions and the mass public. *Journal of Global Security Studies*, vol. 6, n° 2, p. ogz077.

Kovoor-Misra S., Clair J. A., Bettenhausen K. L. (2001). Clarifying the attributes of organizational crises. *Technological Forecasting and Social Change*, vol. 67, n° 1, p. 77–91.

Kulikova O., Heil R., Berg J. van den, Pieters W. (2012). Cyber crisis management: A decision-support framework for disclosing security incident information. In *2012 international conference on cyber security*, p. 103–112.

Lauras M., Truptil S., Benaben F. (2015). Towards a better management of complex emergencies through crisis management meta-modelling. *Disasters*, vol. 39, n° 4, p. 687–714.

Merad M., Ouerdane W., Dechy N. (2011). Expertise and decision-aiding in safety and environment domains: what are the risks? In *Esrel annual conference 2011*, p. 2317–2323.

Merad M., Trump B. D. (2020). *Expertise under scrutiny*. Springer.

Mikolaj J. (2005). Crisis management in security environment. *Komunikácie-vedecké listy Žilinskej univerzity v Žiline*, vol. 7, nº 3, p. 29–33.

Miller H., Griffy-Brown C. (2018). Developing a framework and methodology for assessing cyber risk for business leaders. *Journal of Applied Business & Economics*, vol. 20, nº 3.

Provan K. G., Kenis P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of public administration research and theory*, vol. 18, nº 2, p. 229–252.

Renn O. (2011). The social amplification/attenuation of risk framework: application to climate change. *Wiley Interdisciplinary Reviews: Climate Change*, vol. 2, nº 2, p. 154–169.

Renn O. (2021). New challenges for risk analysis: systemic risks. *Journal of Risk Research*, vol. 24, nº 1, p. 127–133.

Renn O., Klinke A., Van Asselt M. (2011). Coping with complexity, uncertainty and ambiguity in risk governance: a synthesis. *Ambio*, vol. 40, p. 231–246.

Renn O., Lucas K., Haas A., Jaeger C. (2019). Things are different today: the challenge of global systemic risks. *Journal of Risk Research*, vol. 22, nº 4, p. 401–415.

Ros G. (2020). The making of a cyber crash: a conceptual model for systemic risk in the financial sector. *ESRB: Occasional Paper Series*, nº 2020/16.

Schweizer P.-J. (2021). Systemic risks–concepts and challenges for risk governance. *Journal of Risk Research*, vol. 24, nº 1, p. 78–93.

Sherman A. T., DeLatte D., Neary M., Oliva L., Phatak D., Scheponik T. *et al*. (2018). Cybersecurity: Exploring core concepts through six scenarios. *Cryptologia*, vol. 42, nº 4, p. 337–377.

Shrivastava P. (1993). Crisis theory/practice: Towards a sustainable future. *Industrial & Environmental Crisis Quarterly*, vol. 7, nº 1, p. 23–42.

Skagerlund K., Forsblad M., Slovic P., Västfjäll D. (2020). The affect heuristic and risk perception–stability across elicitation methods and individual cognitive abilities. *Frontiers in psychology*, vol. 11, p. 970.

Sommer P., Brown I. (2011). Reducing systemic cybersecurity risk. *Organisation for Economic Cooperation and Development Working Paper No. IFP/WKP/FGS (2011)*, vol. 3.

Trimintzios P., Holfeldt R., Koraeus M., Uckan B., Gavrila R., Makrodimitris G. (2015). *Report on cyber crisis cooperation and management: Comparative study on the cyber crisis management and the general crisis management*.

Van Asselt M. (2000). *Perspectives on uncertainty and risk: the prima approach to decision support*. Springer Science & Business Media.

Van Schaik P., Renaud K., Wilson C., Jansen J., Onibokun J. (2020). Risk as affect: The affect heuristic in cybersecurity. *Computers & Security*, vol. 90, p. 101651.

Wilde G. J. (1998). Risk homeostasis theory: an overview. *Injury prevention*, vol. 4, nº 2, p. 89–91.

Zuccaro G., De Gregorio D., Leone M. F. (2018). Theoretical model for cascading effects analyses. *International journal of disaster risk reduction*, vol. 30, p. 199–215.

# Risques induits par l'intelligence artificielle

## Une approche d'aide à l'identification

**Jacky Akoka[1], Isabelle Comyn-Wattiau[2]**

*1. Laboratoire CEDRIC-CNAM*
*2 Rue Conté, 75003 Paris, France*
*jacky.akoka@lecnam.net*

*2. ESSEC Business School*
*3 Avenue Bernard Hirsch, 95021 Cergy Cedex, France*
*isabelle.wattiau@essec.edu*

*RÉSUMÉ. L'intelligence artificielle est de plus en présente dans la vie des organisations. Elle offre beaucoup d'opportunités pour améliorer la performance de ces organisations. Son utilisation peut les exposer à des risques spécifiques, avec des conséquences potentiellement graves. Pour aider les organisations à repérer ces risques, cet article propose une approche d'aide à l'identification de ceux-ci. L'approche s'appuie sur un modèle conceptuel et trois typologies, respectivement des risques, des techniques d'intelligence artificielle et des processus métiers. Ces typologies sont croisées dans des matrices permettant d'identifier les menaces spécifiques. Deux matrices sont ainsi combinées pour déduire tous les risques potentiellement associés à l'usage de l'intelligence artificielle dans un processus métier. L'approche est illustrée à l'aide des processus métiers de l'assurance.*

*ABSTRACT. Artificial Intelligence is increasingly present in the life of organizations. It offers many opportunities to improve the performance of these organizations. However, its use can expose them to specific risks with potentially serious consequences. This paper proposes an approach helping organizations to identify these risks. This approach is based on a conceptual model and three typologies: risks, artificial intelligence techniques, and business processes. These typologies are cross-referenced in matrices to identify specific threats. Two matrices are combined to derive all the risks potentially associated with the use of artificial intelligence in a business process. The approach is illustrated using insurance business processes.*

*MOTS-CLES : intelligence artificielle, risque, donnée, approche méthodologique, typologie.*

*KEYWORDS: artificial intelligence, risk, data, methodological approach, typology.*

## 1. Introduction

L'intelligence artificielle (IA) est présente dans tous les secteurs d'activité, qu'ils soient de nature industrielle ou des services. Selon Statista, le chiffre d'affaires du marché mondial des technologies de l'IA est estimé à 200 milliards de dollars en 2023 et devrait atteindre plus de 1800 milliards en 2030[1]. Ces technologies ne sont pas sans présenter un certain nombre de risques.

L'IA offre de nombreuses opportunités tant pour améliorer la productivité par l'automatisation des tâches que pour innover dans les produits et les services. L'IA contribue aussi à la prise de décision en fournissant les prédictions nécessaires à cette décision. Elle est présente dans le commerce électronique (personnalisation des achats, assistants intelligents, prévention des fraudes), dans l'éducation (contenus intelligents, assistants vocaux, apprentissage personnalisé), dans l'automobile (aide à la navigation, véhicule autonome, robotique). Elle pénètre aussi le domaine de la santé (détection des maladies, découverte de médicaments), les ressources humaines (repérage des talents), l'agriculture (analyse des données, robotisation). Elle offre aussi des applications créatives dans le domaine artistique ou dans le design. Enfin, sans pouvoir être exhaustif, mentionnons aussi la finance (aide à la gestion de patrimoine, détection des fraudes, analyse des risques).

Pour rendre possibles ces nouvelles applications, l'IA met en œuvre un certain nombre de techniques. Les techniques d'apprentissage automatique fondées sur l'exploitation des données permettent l'acquisition de connaissances. Elles peuvent être supervisées ou non, en milieu fermé ou en interaction avec l'environnement. Elles permettent de transférer une connaissance d'un domaine à l'autre, comme le raisonnement par analogie. La vision assistée par ordinateur reproduit la capacité de perception des images par l'homme. Le traitement de la langue naturelle permet de transférer à la machine les capacités humaines de communication. Plus récemment, l'IA générative a montré la capacité de l'ordinateur à générer des nouveaux contenus (textes, images, vidéos, etc.) par compilation de grands modèles d'informations pré-enregistrées.

Ces opportunités et les techniques qu'elles exigent ne sont pas sans générer leur lot de risques. Certains de ces derniers découlent de l'usage de données dont la qualité n'est ni totale ni connue. Par là-même, l'IA s'appuie sur une information incomplète ou erronée qui peut biaiser son raisonnement. D'autres risques sont de nature juridique dans la mesure où la transparence n'est pas toujours possible. Même transparent, un processus d'intelligence artificielle peut ne pas être conforme aux réglementations liées à l'usage des données tout au long de leur cycle de vie. C'est une révolution qui va aussi engendrer des mutations profondes sur le marché de l'emploi, laissant sur le côté celles et ceux qui voient leurs tâches effectuées plus facilement par des robots fiables, résilients et obéissants. L'IA requiert

[1] https://www.statista.com/topics/3104/artificial-intelligence-ai-worldwide/#topicOverview

l'accumulation de masses de données qu'il est difficile de gérer et contrôler, engendrant des craintes fondées quant au respect de la vie privée.

Parmi tous ces risques, il convient aussi de mentionner le risque à ne pas faire, laissant la concurrence prendre le dessus. C'est la raison pour laquelle toutes les organisations sont concernées par l'anticipation de ces risques. La question de recherche ciblée dans cet article est : comment peut-on identifier les risques liés à l'utilisation de l'IA dans un processus métier ? L'objet de cet article est de proposer une approche permettant de recenser et comparer les risques liés à l'usage de l'intelligence artificielle. L'approche doit être évolutive, généralisable à différentes activités et différents secteurs. Elle doit faciliter la tâche de compréhension du risque et de structuration de la réponse.

Le reste de cet article est organisé comme suit. La section 2 présente un état de l'art sur les techniques de l'IA et les risques associés. La section suivante décrit l'approche en déroulant chacune de ses étapes. Elle est instanciée au domaine de l'assurance. La dernière section est consacrée à la conclusion et aux recherches futures.

## 2. Revue de la littérature

Dans cette section, nous présentons et analysons la littérature relative aux typologies des techniques d'IA et celles relatives aux risques induits.

### *2.1. Sur les typologies de l'intelligence artificielle*

De nombreuses typologies de l'IA sont présentées dans la littérature. Certaines sont fondées sur les capacités. Par exemple, ont été successivement définis les termes Intelligence Artificielle, puis IA Générale (qui effectue toutes les tâches intellectuelles dont l'humain est capable, inclus les tâches de créativité), puis, par réaction, l'IA étroite (restreinte à des tâches spécifiques, par exemple la reconnaissance faciale) et la super IA (qui peut surpasser l'intelligence humaine). Selon (Pereira *et al*., 2023), l'IA est parfois classée selon la faculté humaine qu'elle imite : apprentissage, raisonnement, planification, perception, etc. La classification par capacité proposée par (Schmid *et al.,* 2021) étend et détaille celle par faculté humaine : Sentir, Traiter et comprendre, Agir, Communiquer sont les 4 catégories du premier niveau. Au deuxième niveau, on a 10 capacités, par exemple Natural Language Processing (NLP) et Interaction homme-machine pour la partie Communiquer.

Une deuxième façon de classifier l'IA consiste à prendre en considération les fonctionnalités qu'elle prend en charge, par exemple en quatre catégories : l'IA réactive (sans capacité d'apprentissage), l'IA à mémoire limitée, l'IA à théorie de l'esprit (intégrant les pensées et les émotions), l'IA consciente d'elle-même (dotée de sensibilité) (Hintze, 2016).

Il existe d'autres façons de classifier, par exemple par domaine d'application, par la valeur apportée, par dimension (organisationnelle, humaine, technologique ou système d'information (SI) (Lee *et al.,* 2023), avec des guidelines pour mettre en place l'IA). Ils structurent de cette façon les antécédents, les conséquences, les défis et les guidelines pour implémenter l'IA dans les organisations. Enfin, on peut trouver des classifications selon les techniques mises en œuvre. Si l'on s'en tient à des articles très cités, mentionnons (Borges *et al*., 2021) qui structurent les domaines d'application de l'IA par ce qu'elle apporte (création de valeur) : l'aide à la décision, l'engagement du client ou de l'employé, l'automatisation, les nouveaux produits ou services. Collins *et al*. (2021) répertorient la recherche en Intelligence Artificielle pour en recenser la valeur business et les contributions dans le domaine des SI. Ils proposent une classification de l'IA en : apprentissage automatique (« machine learning »), visionique (computer vision), traitement de la langue naturelle, robotique, systèmes experts, etc. Ce qui différencie ces classifications, c'est leur granularité et/ou le périmètre plus ou moins vaste qu'elles couvrent. Certaines classifications ne couvrent en fait que l'apprentissage automatique, laissant de côté les techniques classiques de raisonnement, tels les systèmes experts. C'est ce type de classification qui nous intéresse dans cet article dans la mesure où les risques sont induits par les techniques mobilisées.

### *2.2. Sur les typologies de risques liés à l'IA*

De nombreuses listes de risques sont aussi présentes dans la littérature. Elles ne sont pas forcément structurées ni exhaustives. Citons toutefois l'article de (McLean *et al*., 2023) qui cible l'IA générale (AGI pour Artificial General Intelligence). Il distingue les catégories de risques suivantes : perte de contrôle par l'humain, développement d'objectifs dangereux, développement d'AGI dangereux, éthique médiocre, risques existentiels. Cette liste est très spécifique, ciblant l'IA générale dont tous les experts disent qu'elle n'est pas encore opérationnelle. Certains articles étudient les risques éthiques qui sont classés selon le principe qu'ils violent : transparence, respect de la vie privée, imputabilité, équité sont les plus mentionnés (Khan *et al*., 2022). Certaines typologies ciblent un domaine, par exemple la santé : Muley *et al.* distinguent les risques liés aux données cliniques, les risques techniques et les risques socio-éthiques (Muley *et al.,* 2023). Dans (Akoka et Comyn-Wattiau, 2022), nous avons proposé une typologie des risques liés aux données, construite avec des praticiens. Elle structure les risques en trois catégories : stratégiques et réputation, légaux et réglementaires, opérationnels. Le groupe de travail AIRS (Artificial Intelligence Risk and Security)[2] propose une classification en quatre catégories : risques relatifs aux données, les attaques contre les systèmes de « machine learning », les risques liés au manque de transparence (incluant les biais), le risque de non-conformité.

Il existe des publications plus mathématiques sur l'évaluation du risque. Citons par exemple (Giudici *et al*., 2024) qui propose quatre indicateurs de risques SAFE

[2] https://www.airsgroup.ai/artificial-intelligence-governance

(pour Sustainability, Accuracy, Fairness, Explainability) et des modèles mathématiques pour les estimer. A notre connaissance, il n'existe pas d'approche structurée pour guider les décideurs dans l'identification du risque associé à l'usage des techniques d'intelligence artificielle. Mentionnons l'approche de Buehler et al. qui proposent de confronter les six catégories de risques (respect de la vie privée, sécurité, honnêteté, transparence et explicabilité, sûreté et performance, risques tiers) et les six contextes business (données, sélection et entraînement de modèle, déploiement et infrastructure, contrats et assurance, légal et réglementaire, organisation et culture) (Buehler *et al.*, 2021). Il manque cependant une opérationnalisation de la matrice à un grain plus fin. Les contextes business ne sont pas directement associés à des processus métiers. Cette publication émane d'un cabinet de conseil qui n'a pas rendu disponible plus de matériel. C'est pour combler cette lacune que nous proposons une approche décrite dans la section suivante de cet article.

## 3. L'approche

Dans cette section, nous décrivons notre approche d'aide à l'identification des risques induits par l'usage de l'IA pour les organisations. Nous nous appuyons sur les sciences de conception (design science) en commençant par décrire les exigences qui ont guidé nos choix dans l'élaboration de la méthode. La deuxième partie décrit le modèle conceptuel sous-jacent qui est le support des différentes étapes qui sont décrites dans la suite.

### *3.1. Les exigences*

Notre question de recherche conduit à la proposition d'une méthodologie. Keller et Binz ont étudié les exigences auxquelles les méthodologies de conception doivent répondre (Keller et Binz, 2009). Nous avons repris et adapté celles qui nous semblaient les plus pertinentes face aux demandes des responsables qui souhaitent une aide pour identifier les risques qu'ils prennent quand ils adoptent l'IA. Ainsi, on retrouve l'exigence *d'utilité* (« usefulness »), de *compréhensibilité*, d'*apprentissage* (learnability). La *spécificité du problème* réside dans le périmètre d'application de la méthode qui doit apporter le guidage pour identifier tous les risques potentiels sans néanmoins fournir l'expertise d'évaluation chiffrée du risque. L'approche doit faciliter la *structuration* de la démarche au moyen d'étapes clairement définies et logiquement enchaînées. Elle doit aussi être *flexible* pour permettre l'évolution des contenus. Constatant l'absence d'aide structurée pour l'identification des risques induits par l'usage de l'intelligence artificielle, nous ciblons une *approche générique* qui englobe tous les types de risques, associés à chaque technique d'IA et prenant en compte tous les processus métiers. Une telle approche doit être *robuste* pour s'appliquer à différents domaines et prendre en charge de façon dynamique les nouvelles menaces au gré de l'évolution des techniques d'IA. Elle doit être *efficace* et *rentable*. Son utilisation répétée doit en faciliter l'enrichissement progressif et permettre une fertilisation croisée entre les domaines auxquels on l'applique.

L'approche doit être *semi-automatique*. En effet, elle ne peut être automatique du fait de sa complexité. Elle ne peut être manuelle du fait de son caractère fastidieux. En revanche, elle doit offrir des guides en ligne qui facilitent l'identification du risque. Elle doit aussi permettre *l'apprentissage* des concepts et des règles qui les régissent. Sa mise en œuvre doit être *facile* du fait qu'elle s'adresse en premier lieu à des décideurs non informaticiens qui sont essentiellement des managers. C'est sur la base de ces exigences que nous avons structuré l'approche décrite ci-après.

### *3.2. Vue d'ensemble de l'approche*

Dans cette partie, nous décrivons successivement le modèle conceptuel sur lequel se fonde l'approche puis ses étapes principales.

#### *3.2.1. Le modèle conceptuel*

Ce modèle rassemble les concepts utilisés dans l'approche. Au niveau le plus élémentaire, on associe un **risque** à une **menace**, par exemple *prioriser le profit aux dépens du bien-être* est un risque associé à une menace de biais. Les risques sont regroupés en **sous-catégories** hiérarchiquement rattachées à des **catégories**. Par exemple, le risque de *prioriser le profit aux dépens du bien-être* relève de la sous-catégorie des *risques éthiques*, elle-même rattachée à la catégorie des *risques de stratégie et réputation*. Un même risque peut être rattaché à plusieurs sous-catégories. Ainsi, le risque de *discriminer via les données* relève de la sous-catégorie *Environnement-Social-Gouvernance (ESG)* de la catégorie *risques de stratégie et réputation*, mais aussi de la sous-catégorie *conformité à la loi* de la catégorie *lois et réglementations*. La menace de biais résulte de l'utilisation de la technique de traitement de la langue naturelle **(technique d'IA)**, qui peut être utilisée pour traiter automatiquement les sinistres en assurance (**processus métier**).

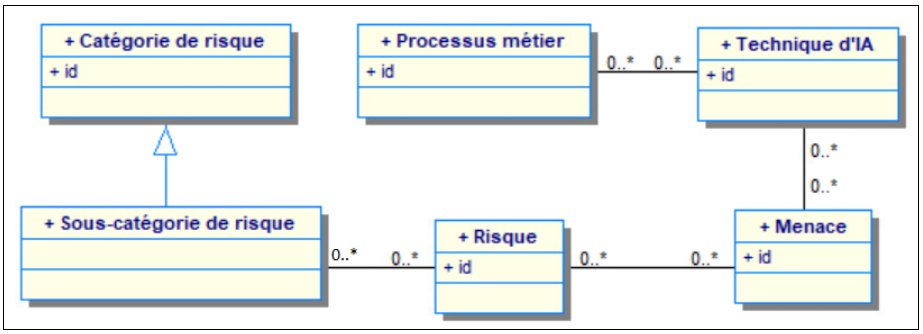



*Figure 1. Le modèle conceptuel sous-jacent*

#### *3.2.2. Les étapes de l'approche*

L'identification des risques liés à l'IA requiert un effort conséquent dans la mesure où il faut considérer, un par un, tous les types de risques et les confronter à tous les usages potentiels de l'IA dans le processus métier étudié. C'est la raison

pour laquelle nous proposons de procéder à une identification systématique qui soit le plus possible généralisable à différents domaines. De plus, cet effort d'identification nécessite différentes compétences (connaissance des processus métiers, compréhension des techniques d'IA et maîtrise des types de risques). En décomposant les étapes de cette identification des risques, on peut atteindre ces deux objectifs de généralisabilité et de répartition par domaine de compétences.

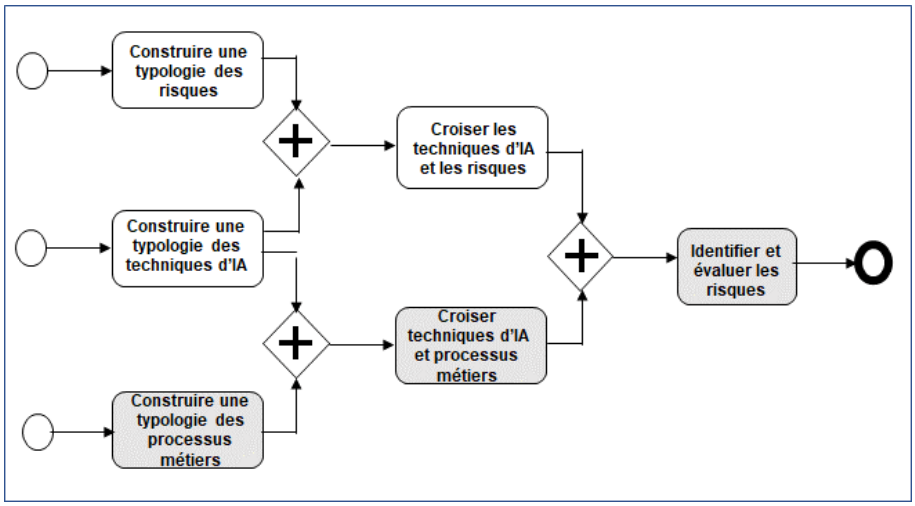



*Figure 2. Les étapes de l'approche*

Notre démarche comporte six étapes (Figure 2). Les tâches sur fond gris sont spécifiques aux processus métiers d'un domaine. Au premier niveau à gauche et en parallèle, on construit trois typologies respectivement des risques, des techniques d'IA et des processus métiers relatifs à un domaine. Ces typologies sont, dans une certaine mesure, indépendantes et peuvent être construites séparément. A noter qu'elles ne peuvent pas être figées parce qu'elles modélisent des domaines très évolutifs. Au deuxième niveau, suivent deux étapes de mise en correspondance (mapping). L'une de ces étapes consiste à croiser les techniques d'IA et les risques potentiels. La seconde met en relation les processus métiers et les techniques d'IA qu'ils peuvent mobiliser. La première est indépendante du secteur d'activité. En la constituant ainsi, on peut ensuite l'appliquer à différents secteurs d'activité. La seconde requiert une connaissance d'un métier ou d'un secteur d'activité. Enfin, au troisième et dernier niveau, l'effort consiste à composer les deux matrices résultats des deux étapes de mise en correspondance. La composition s'effectue sur la dimension commune et permet d'obtenir, *in fine*, une matrice croisant les risques et les processus métiers. L'intérêt de l'approche réside dans la réutilisation de l'effort d'un domaine à un autre : deux typologies génériques et une typologie métier, une matrice de croisement générique et une autre spécifique à un métier.

### *3.3. L'approche générique*

On peut trouver de nombreuses typologies pour chacun des concepts principaux de notre approche. Toutefois, une telle typologie n'est pas pertinente dans tous les

contextes. On cherche ici à définir une approche générale de l'identification des risques liés à l'usage de l'IA, qui soit adaptée ou adaptable à n'importe quel processus métier. Les typologies sont le moyen de définir, au grain adéquat, les concepts à mettre en correspondance.

*3.3.1 Typologie des risques*

Rappelons qu'il n'existe pas de typologie des risques liés à l'IA qui soit reconnue comme standard. Nous avons repris et adapté celle élaborée dans (Akoka et Comyn-Wattiau, 2022) pour évaluer l'impact négatif des données et l'avons enrichie par consultation de la littérature académique et professionnelle. Parmi les trois catégories de risques (stratégie et réputation, légal et réglementaire, opérationnel), la table 1 liste ceux de la première catégorie. Chaque catégorie est ensuite décomposée en sous-catégories. La catégorie *Stratégie et réputation* comprend ainsi les sous-catégories : ESG, Ethique, Confiance et Prise de décision. L'aspect *Légal et réglementaire* est lui-même décomposé en la Conformité aux lois, la Conformité aux réglementations et à la Propriété intellectuelle. Enfin, la catégorie *Risques opérationnels* se décompose selon les trois dimensions Personnes, Processus et Technologies. Nous avons compilé la littérature pour confronter cette hiérarchie aux différents risques liés à l'usage de l'IA et avons pu, dans chaque cas, trouver une correspondance simple ou multiple entre la menace et une ou plusieurs sous-catégories de risques. Dans la table 1, la première colonne décrit la sous-catégorie de risque. La seconde colonne liste les risques relatifs à cette sous-catégorie, expliqués partiellement dans la colonne 3. Pour des raisons d'espace, la table 1 ne recense que les éléments de la catégorie risques stratégiques et de réputation. A titre d'exemple, les données personnelles sont, dans la plupart des pays, soumises à des réglementations ou des lois qui prévoient des sanctions importantes en cas de non-respect. Ainsi la conformité aux réglementations est illustrée par les risques « enfreindre le Règlement Général sur la Protection des Données (RGPD) » et « enfreindre le règlement japonais Protection of Personal Information Act (APPI) ». La typologie initiale était limitée aux risques liés aux données, alors que l'IA englobe tous les risques liés aux données mais génère des risques aussi au niveau des modèles et des systèmes (Schneider *et al*., 2023). Dans la table 1, les « nouveaux » risques sont en italiques. Par exemple, pour les risques ESG, on a identifié au moins cinq risques au lieu de trois, incluant les tensions sociales liées à toutes les craintes pour le futur de l'emploi et l'attention accrue des médias sur le sujet de l'IA. A noter que la construction de la typologie des risques a mobilisé les entités Risque, Sous-catégorie de risque et Catégorie de risque ainsi que leurs relations respectives (Figure 1). Par exemple, les sous-catégories ESG, Ethique, Confiance et Prise de décision appartiennent de manière exclusive à la catégorie Stratégie et réputation. Cela est traduit par la relation d'héritage. Le risque *Disséminer une information erronée* appartient à au moins deux sous-catégories différentes, respectivement ESG et Confiance, ce qui est traduit par la relation « appartient à » de type n:n. Certains risques appartiennent même à des catégories différentes.

*Table 1. Risques stratégiques et de réputation*

| Sous-catégorie | Risque | Explication |
|---|---|---|
| ESG (Environnement-Social-Gouvernance) | Augmenter l'empreinte carbone | Les systèmes d'IA sont coûteux en temps machine |
| | Discriminer via les données | L'IA peut prendre des décisions biaisées ou inexactes et traiter de manière inéquitable certains groupes de personnes |
| | Disséminer une information erronée (sur la gouvernance) | L'IA utilisée par le site web peut générer des informations erronées |
| | *Attirer l'attention des médias* | *L'attention portée par les médias à l'IA peut amplifier les risques de réputation* |
| | *Générer des tensions sociales* | *L'automatisation grâce à l'IA fait craindre des pertes d'emploi* |
| Ethique | Disséminer une information fallacieuse | *L'IA peut prendre des décisions biaisées ou inexactes et traiter de manière inéquitable certains groupes de personnes* |
| | Fournir une information non objective sur les produits | Les systèmes d'IA utilisés pour générer des avis sur les produits peuvent être piratés |
| | *Privilégier le profit au détriment du bien-être des clients* | *Les systèmes d'IA peuvent être influencés par les préjugés personnels des personnes qui les ont créés* |
| Confiance | Disséminer une information erronée | *Dépendance excessive à l'égard de l'IA : un expert peut, face à de nouvelles situations se limiter aux seuls cas auxquels l'IA a également accès* |
| | Subir une perte d'information sensible | Les expériences négatives liées à l'IA (atteintes à la vie privée) peuvent nuire à la réputation |
| | *Prendre des décisions singulières ou hétérogènes* | Les expériences négatives liées à l'IA (décisions biaisées) peuvent nuire à la réputation |
| | *Fournir des informations non étayées* | *Les modèles d'IA sont difficiles à interpréter et compliquent l'explication des décisions aux clients et autres interlocuteurs* |
| Prise de décision | Prendre une décision sur la base de données erronées | Prendre des décisions erronées, entraînant des pertes financières et/ou la désaffection des clients |
| | Prendre une décision sur la base de données obsolètes | |
| | *Prendre une décision convenue* | *Certaines techniques d'IA ne font pas preuve de créativité (exemple : les systèmes experts)* |

*3.3.2. Typologie des techniques d'IA*

Cet article fait suite à une présentation faite devant un public de professionnels et a permis d'aboutir à cette typologie. Ainsi, face à toutes les typologies des techniques d'IA disponibles, nous avons retenu une typologie très simple à un niveau qui nous semble, à ce stade, de nature à permettre une confrontation, d'une part, aux processus métiers susceptibles d'y recourir et, d'autre part, aux risques que ces techniques peuvent engendrer. Cette typologie comprend les six catégories suivantes :

– l'apprentissage automatique, incluant l'apprentissage profond (*deep learning*), l'apprentissage supervisé ou non, l'apprentissage par renforcement (*reinforcement learning*), etc.

– le traitement de la langue naturelle et la fouille de texte (*text mining*),

– la visionique (*computer vision*),

– la robotique ou automatisation,

– les systèmes experts (incluant les systèmes à base de règles, la représentation des connaissances, le raisonnement dans l'incertain, etc.),

– l'IA générative.

Cette typologie n'est pas une partition : certaines autres techniques sont moins répandues comme les algorithmes génétiques et n'y figurent pas, d'autres techniques se retrouvent dans plusieurs catégories. Ainsi certaines techniques de NLP font partie de l'IA générative et/ou font appel aux techniques d'apprentissage. Cette typologie mobilise l'entité Technique d'IA du modèle conceptuel (Figure 1). A noter que notre choix de typologie couvre plus que l'apprentissage automatique sans toutefois être exhaustif, notamment pour certaines techniques plus rares comme les algorithmes génétiques.

Une fois les deux typologies construites, l'approche consiste à les croiser pour faciliter l'identification des risques.

*3.3.3. Mise en correspondance des techniques d'IA et des risques*

L'objectif de cette étape est de remplir la matrice dont les colonnes sont les techniques d'IA (paragraphe 3.3.1) et les lignes sont les risques associés (paragraphe 3.3.2) (Table 2). Cette étape a été réalisée en analysant les informations disponibles sur les sites et la littérature scientifique traitant du sujet. L'avantage de l'approche est de pouvoir la bâtir indépendamment du domaine d'application.

Le processus de mise en correspondance vise à identifier le type de menace encourue lors de l'usage de la technique d'IA en s'aidant de la liste de tous les risques potentiels. Pour matérialiser ce risque, on caractérise son existence par la menace sous-jacente. Par exemple, le risque de discrimination via les données représente une menace qualifiée de biais lors du recours au machine learning, ou au traitement de la langue naturelle ou à la visionique. Un autre exemple est celui d'un robot qui utilise des informations erronées ou obsolètes (risque « prendre des décisions sur la base de données erronées ou obsolètes ») et peut, par là-même générer une menace de sécurité physique s'il est en interaction avec des personnes

(véhicule autonome par exemple). Cette étape matérialise l'entité Menace reliée d'une part à la technique d'IA et d'autre part à l'entité Risque (Figure 1).

*Table 2. Correspondance entre techniques d'IA et risques*

| Technique d'IA / Risque | Machine Learning | NLP | Visionique | IA générative | Systèmes experts | Robotique |
|---|---|---|---|---|---|---|
| Augmenter l'empreinte carbone | Impact Environnemental | | | | | |
| Discriminer via les données | Biais | | | | | |
| Disséminer une information erronée sur la gouvernance | | | | Altération de la réalité | | |
| Attirer l'attention des médias | Impact Gouvernance | | | | | |
| Générer des tensions sociales | | | | Impact Social | | |
| **Risque Ethique** | | | | | | |
| Disséminer une information fallacieuse | Biais | | | | | |
| Fournir une information non objective sur les produits | Violation de l'intégrité | | | Violation de l'intégrité | | |
| Privilégier le profit au détriment du bien-être des clients | Biais | | | | | |
| **Risque Confiance** | | | | | | |
| Disséminer une information erronée | | | | Altération de la réalité | | |
| Subir une perte d'information sensible | | Violation de la vie privée ; perte de confidentialité | | | | |
| Prendre des décisions singulières ou hétérogènes | Biais | | | | Mauvaise déduction | |
| Fournir des informations non étayées | Opacité | | | Opacité | | |
| **Risque prise de décision** | | | | | | |
| Prendre une décision sur la base de données erronées | Mauvaise décision | | | Biais, opacité, mauvaise décision | Mauvaise décision | Insécurité physique |
| Prendre une décision sur la base de données obsolètes | | | | | | |
| Prendre une décision convenue | | | | | Manque de créativité | |

Pour des raisons d'espace, la table 2 ne contient que les risques stratégiques et de réputation. Bien entendu, cette table n'est pas exhaustive, le nombre et la nature des menaces étant très évolutifs. Dans l'avenir, on pourrait utiliser cette matrice comme structure pour une base de connaissances qui accumulerait les événements publiés sur ces menaces.

### *3.4. Application de l'approche aux processus métier de l'assurance*

Il nous faut en premier lieu construire une typologie des processus métiers de l'assurance.

#### *3.4.1. Typologie des processus métiers de l'assurance*

Une organisation qui souhaite identifier les risques qu'elle court si elle met en place l'IA doit, au préalable, lister les processus métiers qui sont concernés. A titre d'illustration, nous avons élaboré une typologie des processus métiers du domaine de l'assurance en nous inspirant de la littérature professionnelle[3]. L'examen de celle-ci permet de recenser les processus métiers bénéficiaires de l'IA[4]. En combinant ce recensement avec la chaîne de valeur de Porter (Robben, 2014), nous avons obtenu la liste suivante : souscription de contrat, fixation des prix et évaluation des risques, service client et fidélisation, traitement des sinistres, atténuation des risques et prévention des pertes, administration des polices d'assurance, mise en conformité réglementaire, modélisation du risque et réassurance, rétention des clients, détection et prévention de la fraude.

Certains processus sont très spécifiques à l'assurance mais d'autres sont plus standards, par exemple le service client et fidélisation ou la conformité réglementaire. Cette typologie mobilise l'entité Processus métier du modèle conceptuel (Figure 1).

#### *3.4.2. Mise en correspondance des techniques d'IA et des processus métiers*

Au cours de cette étape, nous générons la matrice résultant de la mise en correspondance des processus métier, ici l'assurance, et des techniques d'IA (Table 3). Par exemple, le processus Souscription de contrat peut bénéficier de techniques de langage naturel afin de générer le texte du contrat par composition d'articles extraits d'une base de contrats-type. Il peut bénéficier aussi du machine learning qui analyse l'acceptabilité du client fondée sur ses données personnelles (classe d'âge, historique d'accidents, état de santé, revenu, etc.). Un autre exemple propre à l'assurance est celui du traitement des sinistres qui peut bénéficier de l'apport du traitement du langage naturel pour analyser les documents envoyés par le client, de robotique pour capter les images décrivant le sinistre et de visionique pour analyser ces images. Cette phase de l'approche mobilise la relation Utilise entre les entités Processus métier et Technique d'IA.

---

[3] https://acpr.banque-france.fr/sites/default/files/medias/documents/20220114_as132_transfo_numerique_assurance.pdf

[4] https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/implementing-generative-ai-with-speed-and-safety

*Table 3. Correspondance entre techniques d'IA et processus de l'assurance*

| IA<br>Processus | Machine Learning | NLP | Visionique | IA générative | Systèmes experts | Robo-tique |
|---|---|---|---|---|---|---|
| **Souscription de contrat** | Analyse de données | Extraction de texte | | | | |
| **Fixation des prix et évaluation des risques** | Modélisation des facteurs de risques ; analyse des patterns | | | Reconnaissance de "patterns" | Estimation du risque en fonction du profil | |
| **Service client et fidélisation** | Systèmes de recommandation | Analyse de sentiments ; chatbots et assistants virtuels | | Assistants virtuels ; expérience personnalisée | | |
| **Traitement des sinistres** | | Analyse de documents | Analyse d'image | | | Capture d'image |
| **Atténuation des risques et prévention des pertes** | Méthodes ensemble ; analyse prédictive | | | | | |
| **Administration des polices d'assurance** | | | | | | Automatisation robotisée des processus |
| **Conformité réglementaire** | | Vérification de document | | | Vérification via règles | |
| **Modélisation du risque et réassurance** | Analytique avancée | | | | | |
| **Rétention des clients** | Prédiction du "churn" | Communication personnalisée | | Expérience personnalisée | | |
| **Détection et prévention de la fraude** | Détection d'anomalies ; catégorisation des sinistres | | | | Identification de schémas de fraude | |

### *3.4.3. Identification des risques*

Le but de cette phase est de fournir au décideur une matrice résultant de la composition des deux précédentes grâce aux techniques d'IA qui sont la dimension commune (Table 4). Elle comprend en ligne tous les risques triés par catégorie et en colonne les processus métiers (ici l'assurance). La table 4 n'est qu'un extrait de trois processus croisés avec six risques. Le contenu de chaque cellule a été obtenu par réinterprétation de la composition des deux matrices précédentes. Par exemple, en partant du risque « Générer des tensions sociales », la matrice de la table 2 met en

avant l'impact social de la robotique, de l'IA générative et des systèmes experts. La matrice de la table 3 repère ces techniques dans le processus de traitement des sinistres (capture d'image par robot) et dans le processus d'administration des polices d'assurance lui aussi impacté par l'automatisation par robots logiciels. La matrice de la table 4 résulte d'une réinterprétation du résultat obtenu par composition.

*Table 4. Identification des risques pour les processus d'assurance*

| **Processus** / **Risque** | **Souscription de contrat** | **Traitement des sinistres** | **Administration des polices d'assurance** |
|---|---|---|---|
| **Risque ESG** | | | |
| **Discriminer via les données** | Biais dans l'analyse de données ; biais dans la composition automatique de texte | Analyse biaisée des documents et des images | |
| **Générer des tensions sociales** | | Tensions sociales liées au remplacement des experts d'assurance par la robotique | Tensions sociales liées au remplacement des gestionnaires de contrat par des robots |
| **Risque éthique** | | | |
| **Disséminer une information fallacieuse** | Biais dans l'analyse de données ; biais dans l'extraction de texte | Biais dans l'analyse des documents et des images du sinistre | |
| **Privilégier le profit aux dépens du bien-être** | Biais dans l'analyse de données | | |
| **Risque confiance** | | | |
| **Subir une perte d'information sensible** | | Divulgation d'informations privées dans l'analyse des documents et des images du sinistre | |
| **Prendre des décisions singulières ou hétérogènes** | Biais dans l'analyse des données | Mauvaise conclusion tirée de l'analyse des documents et des images | |
| **Risque prise de décision** | | | |
| **Prendre une décision sur la base de données obsolètes** | Non applicable | | |
| **Prendre une décision convenue** | | | |

Cette matrice peut servir à différents décideurs. Le responsable du processus métier peut utiliser le résultat pour anticiper les risques et les menaces pouvant résulter de l'utilisation de l'IA dans son champ de responsabilité. Entre différentes techniques IA, on pourrait enrichir les matrices en hiérarchisant chacun des risques. Quant au gestionnaire des risques, il se voit faciliter l'audit et la veille en repérant les techniques qui sont potentiellement génératrices de nouveaux risques et les entités concernées. Enfin, le data scientist, en charge de certaines techniques d'IA, est garant de la complétude des matrices des Tables 2 et 3 et peut assister les responsables de processus dans leurs choix de techniques, compte-tenu des risques encourus. Cette matrice peut être utilisée à différentes fins, notamment pour structurer la jurisprudence faisant état des dysfonctionnements rendus publics sur ces risques.

## 4. Conclusion et recherche future

Face à l'explosion de l'offre de solutions logicielles à base d'IA, les entreprises et organisations sont conscientes des opportunités mais aussi des risques. Pour ce second point, elles ne disposent pas facilement d'aide à leur identification. Nous avons présenté, dans cet article, une approche qui capitalise sur des typologies de risques, de techniques IA et de processus métiers pour faciliter l'identification de tous les risques stratégiques, juridiques et opérationnels encourus du fait de l'utilisation des techniques d'IA. L'approche fournit un moyen systématique d'analyser les conséquences de l'introduction d'une technique d'IA dans un processus. L'originalité principale est de composer une matrice générique IA*risques et une matrice spécifique IA*processus métiers. La première, une fois construite et validée, peut être enrichie par des experts du domaine de l'IA, accompagnés par des spécialistes du risque. La seconde doit être mise en place dès qu'on considère un nouveau domaine d'activité. Elle requiert les experts métiers pour cartographier les processus et peut s'appuyer sur la presse professionnelle qui met en avant les opportunités de l'IA, même sans en mentionner les risques, lesquels sont ensuite obtenus au moyen de l'autre matrice. A l'image de l'effort que les assureurs déploient pour couvrir le risque cyber, l'approche peut aussi être utilisée par eux pour proposer une offre de service à leurs clients en vue de couvrir l'usage de l'IA.

Dans la recherche future, nous prévoyons de compléter la batterie de matrices avec les outils d'évaluation mathématique des risques ainsi que les moyens d'atténuation. Une autre extension consistera à assortir les risques d'une échelle permettant de les prioriser. Pour mener plus avant la validation, nous prévoyons aussi de tester l'approche dans d'autres domaines, par exemple la logistique. Les matrices présentées dans l'article sont encore en construction et requièrent des experts de différents profils pour les valider et les faire évoluer. Enfin, le développement d'un outil semi-automatique d'aide à la décision pour faciliter le parcours des matrices et leur composition est à l'étude.

*Remerciements. Les auteurs remercient les partenaires de la Chaire ESSEC Stratégie et gouvernance de l'information où cette recherche a été menée.*

## Bibliographie

Akoka J., Comyn-Wattiau I. (2022). Evaluation de la valeur des données - Modèle et méthode. *40ème congrès INFORSID*, Association Inforsid, Dijon, France. p.163-178.

Borges A., Laurindo J., Spinola M., Gonçalves R. Mattos C. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions, *International Journal of Information Management*, vol. 57, 102225.

Buehler K., Dooley R., Grennan L., & Singla,A. (2021). Getting to know—and manage—your biggest AI risks. *Mckinsey and Company*.

Collins C., Dennehy D., Conboy K., Mikalef P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda, *International Journal of Information Management*, vol. 60, 102383, ISSN 0268-4012.

Giudici P., Centurelli M., Turchetta S. (2024). Artificial Intelligence risk measurement, *Expert Systems with Applications*, vol. 235, 121220.

Hintze, A. (2016) Understanding the Four Types of AI, from Reactive Robots to Self-Aware Beings. *The Conversation*.

Keller A., Binz H. (2009). Requirements on engineering design methodologies. In *DS 58-2: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 2, Design Theory and Research Methodology, Palo Alto, CA, USA*.

Khan A., Badshah S., Liang P., Waseem M., Khan B., Ahmad A., … & Akbar M. A. (2022). Ethics of AI: A systematic literature review of principles and challenges. In *Proceedings of the 26th Intl. Conf. on Evaluation and Assessment in Software Engineering*, p. 383-392.

Lee M., Scheepers H., Lui A., Ngai E. (2023). The implementation of artificial intelligence in organizations: A systematic literature review, *Information & Management*, vol. 60, n°5,103816, ISSN 0378-7206.

McLean S., Read G., Thompson J., Baber C., Stanton N., Salmon P. (2023). The risks associated with Artificial General Intelligence: A systematic review, *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 35, n°5, p. 649-663.

Muley A., Muzumdar P., Kurian G., & Basyal G. (2023). Risk of AI in healthcare: A comprehensive literature review and study framework. *arXiv preprint arXiv:2309.14530*.

Pereira V., Hadjielias E., Christofi M., Vrontis D. (2023). A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective, *Human Resource Management Review*, vol. 33, n°1, 100857, ISSN 1053-4822.

Robben X. (2014). *La chaîne de valeur de Porter : Identifier la création de valeur*. 50 Minutes.

Schmid T., Hildesheim W., Holoyad, T. *et al.* (2021). The AI Methods, Capabilities and Criticality Grid. *Künstl Intell,* vol. 35, p. 425–440.

Schneider J., Abraham R., Meske C., Vom Brocke J. (2023) Artificial Intelligence Governance For Businesses, *Information Systems Management*, vol. 40 n°3, p. 229-249.

# Sécurité dans les SI & *social engineering* - un état des lieux

**Florence Sèdes[1], Jonathan Degrace[2]**

*1. sedes@irit.fr*
*IRIT - Université Toulouse 3 Paul Sabatier*
*118 Route de Narbonne*
*31062 Toulouse cedex 9*
*2. jonathan.degrace@medes.fr*
*MEDES - Clinique Spatiale*
*2 avenue de l'aérodrome de Montaudran*
*31 400 Toulouse*

*RESUME. De grandes transformations liées aux technologies de l'information touchent les Systèmes d'Information (SI) qui soutiennent les processus métier des organisations ainsi que leurs acteurs. Le déploiement dans un environnement complexe concernant des données sensibles, massives et hétérogènes génère des risques aux impacts juridiques, soci(ét)aux et financiers. Ce contexte de transition et d'ouverture rend la sécurité de ces SI centrale dans les préoccupations des organisations. La numérisation de tous les processus et l'ouverture aux dispositifs IdO (Internet des Objets) a favorisé l'apparition d'une nouvelle forme de criminalité : la cybercriminalité.*
*Ce terme recouvre des actions «malicieuses» (malveillantes) dont la majorité sont désormais perpétrées à l'aide de stratégies de social engineering, phénomène permettant une exploitation combinée des vulnérabilités «humaines» et des outils numériques. La «maliciosité» de telles attaques réside dans le fait qu'elles transforment les utilisateurs en facilitateurs des cyberattaques, au point d'en être perçus comme le «maillon faible» de la cybersécurité. Les politiques de déploiement s'avérant insuffisantes, il est nécessaire de réfléchir à des étapes amont : savoir anticiper, analyser signaux faibles et outliers, détecter précocement et réagir promptement sont des questions prioritaires nécessitant une approche axée sur la prévention et la coopération. Dans cet état des lieux, nous proposons un travail de synthèse de la littérature et des pratiques professionnelles à ce sujet.*

*ABSTRACT. Major transformations related to information technologies affect Information Systems (IS) that support the business processes of organizations and their actors. Deployment in a complex environment involving sensitive, massive and heterogeneous data generates risks with legal, social and financial impacts. This context of transition and openness makes the*

*security of these IS central to the concerns of organizations. The digitization of all processes and the opening to IoT devices (Internet of Things) has fostered the emergence of a new form of crime, i.e. cybercrime.*

*This generic term covers a number of malicious acts, the majority of which are now perpetrated using social engineering strategies, a phenomenon enabling a combined exploitation of «human» vulnerabilities and digital tools. The maliciousness of such attacks lies in the fact that they turn users into facilitators of cyber-attacks, to the point of being perceived as the «weak link» of cybersecurity.*

*As deployment policies prove insufficient, it is necessary to think about upstream steps: knowing how to anticipate, identifying weak signals and outliers, detect early and react quickly to computer crime are therefore priority issues requiring a prevention and cooperation approach.*

*In this overview, we propose a synthesis of literature and professional practices on this subject.*

*Mots-clés* : *social engineering - cybercriminalité - attaques - prévention.*

*KEYWORDS: social engineering - cybercrime - attacks – prevention.*

## 1. Introduction

De grandes transformations liées aux technologies de l'information touchent les Systèmes d'Information (SI) qui soutiennent les processus métier des organisations ainsi que leurs acteurs. Le déploiement dans un environnement complexe concernant des données sensibles, massives et hétérogènes génère des risques aux impacts juridiques, sociaux, et financiers. Ce contexte de transition et d'ouverture rend la sécurité de ces SI centrale dans les préoccupations des organisations. La numérisation de tous les processus et l'ouverture aux dispositifs IdO (Internet des Objets / Internet of Things (IoT)) a favorisé l'apparition d'une nouvelle forme de criminalité : la cybercriminalité.

Ce terme générique recouvre un certain nombre d'actes « malicieux » (malveillants) dont la majorité sont désormais perpétrés à l'aide de stratégies de *social engineering* (le terme d'ingénierie sociale en français ne recouvre pas les mêmes acceptions), phénomène permettant une exploitation combinée des vulnérabilités « humaines » et des outils numériques. La « maliciosité » de telles attaques réside dans le fait qu'elles transforment les utilisateurs en facilitateurs des cyberattaques, au point d'en être perçus comme le « maillon faible » de la cybersécurité.

Les particuliers, les entreprises, les institutions et les États sont confrontés au défi de trouver une réponse à ces atteintes. Néanmoins, les moyens juridiques, techniques, économiques et culturels mis en place sont encore insuffisants : loin d'être éradiquée, l'utilisation du *social engineering* à des fins illicites poursuit son essor. Les éléments factuels illustrent ce constat : [1] relève que 74 % des failles de sécurité incluent l'élément humain, que 50 % des attaques de *social engineering* sont de la compromission d'email professionnel et que la cible numéro deux des cyberattaques sont les individus.

Les domaines liés à ces problématiques, comme les aspects juridiques du *social engineering* et les sanctions encourues, ou le rôle des États et des organisations internationales dans la lutte contre la cybercriminalité ne seront pas envisagés dans cette synthèse. L'objectif de ce papier est de situer l'état des lieux des connaissances actuelles concernant le fonctionnement et la prévention du *social engineering* :

- À quel point est-il impliqué dans les incidents en cybersécurité ?
- Quels sont ses vecteurs d'attaques ? Comment fonctionnent-ils ?
- Existe-t-il des mesures de prévention efficaces ?

## 2. Définition du *social engineering*

Historiquement, le terme de *social engineering* vient des sciences politiques. Apparu sous la plume de l'économiste britannique John Gray en 1842, dans son ouvrage, "An efficient Remedy for the distress of Nations", il désignait à l'origine les experts en charge des questions politiques et sociales, par exemple la résolution de la grande

famine irlandaise (1845-1852) qui allait sévir quelques années plus tard. Il sera plus tard usité en sciences sociales, e.g. par J. M. Hatfield 2018 [22] et Slade, John A. 1929 [23]. Son entrée dans le monde cyber se fait dans les années 50 avec le phénomène des “phone phreaking” et les opératrices manuelles, dans le but de détourner le fonctionnement des lignes téléphoniques afin d'accéder à des services spéciaux et / ou de ne pas payer les communications.

Le *social engineering*, encore appelé fraude psychologique ou piratage psychologique, désigne l’utilisation des techniques d’escroqueries historiques adaptées au monde du numérique : on y retrouve des techniques comme la recherche d’informations sur la cible, le mensonge, l’utilisation d’une fausse identité, la tromperie ou la manipulation.

Le but est d’acquérir la confiance de l’autre afin de lui faire réaliser une action frauduleuse comme l’accès à des informations, des biens, des services ou des lieux sécurisés (voir un exemple historique avec le Tupolev, “cousin” du Concorde, “Tupolev-Tu-144-l-espionnage-industriel-au-cœur-de-la-guerre-froide” [15]).

Dans le cadre de cet état de l’art, notre focus se fera sur le *social engineering* impliqué dans les cyberattaques.

Comme évoqué précédemment, 74 % des violations de données incluent un élément humain : vol d’identité, mésusage des accès à privilèges, mauvaises configurations ou *social engineering*, les classant dans le top 3 des causes de violations de données dans 8 secteurs sur 9, et numéro 2 sur la scène internationale [1].

Dans les seules attaques de *social engineering*, la compromission d’email professionnel, à l’aide de technique de phishing, a augmenté de près de 50 % entre 2022 et 2023. Le *phishing*, et ses dérivés (*spear-phishing, vishing, smishing,* etc. cf. 3.1 ci-après), représentent à eux seuls 44 % des attaques en 2023, pour un coût médian de 50 000$ par attaque. C’est actuellement le principal vecteur d’attaque. Parmi ces attaques, 83 % sont des acteurs extérieurs à l'organisation, dont le but est purement lucratif.

À la vue des chiffres indiqués par le rapport 2023 de Verizon [1], on peut ne peut que s’accorder sur l’importance du facteur humain dans la prévention des risques en cybersécurité. En effet, les *social engineers* ne font que profiter de certains des fonctionnements de l’être humain, couplés à l’utilisation des moyens numériques et des techniques de manipulation, comme effet de levier pour leurs attaques (Wang et al. 2021 [2]).

## 3. Les attaques par *social engineering*

### *3.1 Différentes formes d’attaques*

Les attaques peuvent employer, ou non, des moyens techniques comme les emails, les sms, les appels, etc. [30].

Les attaques les plus courantes sont le *phishing* et le *spear phishing* par mail [1]. Le *spear phishing* est un *phishing* personnalisé pour la cible. Nous trouvons certains de ses "dérivés" dans les *smishing* (*phishing* via SMS) ou le *vishing* (*phishing* vocal via message audio ou appel direct). Ces attaques entrent dans la catégorie du *harpooning*, ou hameçonnage en français. Elles consistent à envoyer un message, sous un faux prétexte, pour tromper les victimes afin de leur faire commettre une action qui leur sera préjudiciable, tel un mail urgent de l'école de leurs enfants par exemple, qui leur fait ouvrir un document qui cacherait un ransomware *(phishing / spear-phishing)*. L'envoi d'un sms incitant à ouvrir un lien menant à un faux site de leur banque dans le but d'en dérober les identifiants et mot de passe *(smishing)*, ou un appel, suite à de prétendus prélèvements frauduleux, leur demandant de remettre leurs cartes avec les codes confidentiels à un pseudo-coursier prétendument mandaté par le conseiller bancaire *(vishing*) constituent des stratégies bien rodées. De telles attaques peuvent également avoir lieu à partir d'autres supports comme les réseaux sociaux [10] [24] [31].

À cela, nous pouvons ajouter les attaques physiques en face à face ou à l'aide de *tailgating*, technique consistant à "coller" quelqu'un afin de passer un portique sécurisé par badge par exemple. Un attaquant déguisé en employé pourrait utiliser cette technique dans le but de pénétrer un bâtiment avant de se fondre dans la masse et d'essayer de mener à bien ses actes illicites. Party et Rajendran 2019 [11] présentent le *shoulder surfing* qui consiste à observer par-dessus l'épaule de la victime quand cette dernière saisit des données. On peut facilement imaginer profiter d'un voyage en train ou d'une séance de travail à la terrasse d'un café de notre victime, pour lire ou photographier un écran à l'insu de son propriétaire. La fouille des poubelles de notre cible reste aussi une forme d'attaque de *social engineering*, que cela soit pour trouver des informations pour une attaque ciblée (*spear-phishing*) ou récupérer sur des supports non détruits des informations, pour elle/lui sans importance, mais qui peuvent s'avérer précieuses (dates anniversaires, nom d'animaux, etc. souvent à la base de la création de passwords).

Il existe également des attaques plus techniques et moins directes, comme l'attaque du point d'eau. Cette dernière consiste en la compromission d'un point de rencontre (ex: site web) fréquemment visité par les victimes, comme peuvent parfois le faire des prédateurs avec leurs proies. Elle permet d'atteindre des victimes de façon indirecte par rebond lorsque l'attaque directe est trop complexe ou risquée (S. Kaushalya, R. Randeniya, and A. Liyanage, 2018 [12]).

L'Open Source INTelligence (OSINT) est régulièrement utilisée dans la préparation des attaques (Wang et Al, 2020 [13]). L'OSINT consiste en la recherche d'information à partir de sources ouvertes légalement accessibles, sur internet ou ailleurs. Ces informations peuvent provenir des réseaux sociaux, des registres légaux ou encore de différents médias.

L'arrivée des différentes formes de l'Intelligence Artificielle (IA) et du Machine Learning (ML) permet d'amplifier les effets des attaques de *social engineering*. Ainsi, on assiste à une quasi-industrialisation de la recherche d'informations et des attaques personnalisées de type *spear phishing*. Nous assistons également à l'arrivée d'IA "d'attaque" capables de s'autocorriger afin de s'adapter de plus en plus finement à leurs interlocuteurs pour les tromper (Mouton et al. 2014 [16], Schmitt et Flechais 2023 [17]).

Une liste non exhaustive de différentes méthodes d'attaques peut être trouvée dans les publications de Yasin et al. 2019 [10], Party et Rajendran 2019 [11] et S. Kaushalya, R. Randeniya, and A. Liyanage, 2018 [12] et infosecawerness [28].

### *3.2 Quelles sont les failles humaines et comment sont-elles utilisées ?*

Pour la réussite des différentes techniques énoncées, les *social engineers* s'appuient sur différents aspects du fonctionnement humain, parmi lesquels nous retrouvons certains mécanismes cognitifs, les biais des cibles, les besoins sociaux, les stéréotypes, les heuristiques de pensées et les émotions des individus (Wang et al 2021 [2], Laurent Bègue et Olivier Desrichard 2013 [14] et Yasin et al. 2019 [10]).

Les *social engineers* ne sont bien évidemment eux-mêmes pas à l'abri de ces mêmes biais et autres stéréotypes. On peut citer à ce titre le stéréotype de genre qui induit un biais qui fait que les femmes, réputées moins versées en technologie numérique, sont plus ciblées que les hommes [26].

Ils peuvent aussi compter sur différentes techniques de manipulations ou d'influence pour les aider dans leurs tâches.

Tout cela a pour effet de créer des schémas globaux de potentielles vulnérabilités exploitables en *social engineering* : par exemple, nous serons plus à même d'aider une personne plus proche ou qui nous est physiquement agréable, par biais de halo. D'autres biais cognitifs sont mobilisés, comme le biais de conformité (tendance à penser et agir comme les autres le font), le biais d'ancrage (s'en tenir au premier élément d'information entendu comme référence), l'excès de confiance (tendance à surestimer ses capacités), ... À partir de cette approche, le spécialiste en *social engineering* peut déployer un certain nombre d'outils, que cela soit lors d'une attaque physique ou « virtuelle », à l'aide de l'un des moyens de communication à sa disposition.

De surcroît, lorsque certains traits de personnalités (au sens du *big five*[1]) sont dominants, ils augmentent le risque de tomber dans le piège d'une attaque par *social engineering* [25].

1. https://fr.wikipedia.org/wiki/Mod%C3%A8le_des_Big_Five_(psychologie)

Il est ainsi facile d'imaginer qu'un *social engineer* joue, par exemple, sur le biais de préférence endogroupe [14] à l'aide d'une fausse identité. Ainsi, il peut se rapprocher de sa cible et lui soutirer des informations, des accès ou des biens.

Nous présentons dans le tableau ci-dessous un échantillon non exhaustif des biais, techniques d'influence et de manipulation, que pourrait rencontrer les cibles, lors d'une attaque de *social engineering*. Pour plus d'information, on peut se reporter à Wang et al 2021 [2], Yasin et al. 2019 [10], Caldini 2014 [34] et R-B Joules et J-L Beauvois 2022 [35].

| Biais (cognitif) de la cible | Fonctionnement | Conséquences éventuelles |
|---|---|---|
| Biais de confirmation d'hypothèse | Tendance à rechercher des informations qui confirment nos croyances existantes et à ignorer les informations qui les contredisent : on filtre et on trie les informations selon qu'elles correspondent à nos attentes ou non. | Risque de divulgation d'informations, d'accès non autorisé ou d'action frauduleuse. |
| Aversion à la perte | L'aversion à la perte implique que les individus sont plus sensibles aux perspectives de pertes qu'à celles associées aux gains. | Cela peut mettre la victime sous pression et lui faire prendre un risque inconsidéré dans une attaque de type "Appel du président". |
| Effet Lake Wobegon (ou biais d'auto-complaisance) | L'effet Lake Wobegon traduit la tendance (inconsciente) à penser que nous sommes bien meilleurs que nous ne le sommes en réalité. | La victime pourrait surestimer sa capacité à se défendre face aux différentes formes d'attaques. |
| **Technique d'influence** | **Fonctionnement** | **Conséquence éventuelle** |
| Sentiment de proximité par appartenance | Sentiment de proximité supposé envers un individu qui augmente les probabilités de l'aider. | Risque de divulgation d'informations, d'accès non autorisé ou d'action frauduleuse. |

| | | |
|---|---|---|
| Attractivité physique ou empathique | Sentiment d'attractivité envers un individu qui augmente les probabilités de l'aider. | Risque de divulgation d'informations, d'accès non autorisé ou d'action frauduleuse. |
| Conformisme aux figures d'autorités | Céder plus facilement aux figures d'autorités par normes sociales. | Risque de divulgation d'informations, d'accès non autorisé ou d'action frauduleuse. |
| Technique utilisée | Fonctionnement | Conséquences éventuelles |
| Distraction | Créer une distraction modérée dans le but de surcharger mentalement et diminuer la capacité de réflexion de la cible. | Risque de divulgation d'informations, d'accès non autorisés ou d'action frauduleuse. |
| Fake ID | Utilisation d'une fausse identité sociale dans le but de paraître plus proche de la cible et obtenir son aide. | Tromperie sur l'identité de l'attaquant pouvant mener à un risque de divulgation d'informations, d'accès non autorisés ou d'action frauduleuse. |
| Usage des symboles d'autorités | Utilisation des symboles d'autorités reconnus (ex : blouse, façon de s'exprimer) pour appuyer l'utilisation du fake ID. | Comme le fake ID, avec une pression accrue pouvant mener à un risque de divulgation d'informations, d'accès non autorisés ou d'action frauduleuse plus important. |

### *3.3 Mise en œuvre*

D'après Mouton et al. 2014 [16], une attaque par *social engineering* se décompose en six phases. Le nombre d'étapes et le déroulement pourront varier selon le type d'attaque (physique ou *phishing* par exemple). Nous pouvons les détailler de la manière suivante.

Phase 1 : Formulation de l'attaque :
On détermine quel est le but de cette attaque.
Phase 2 : Récolte d'information :
Analyses des sources d'informations disponibles et de leurs utilisations possibles.
Récolte de l'information
Phase 3 : Préparation de l'attaque
À l'aide des informations récoltées, on vient définir la méthode, éventuellement qui exécute l'attaque et selon quel scénario.
Phase 4 : Développement de la relation de confiance :
Début de l'attaque, tentative de mise en confiance de la cible et de consolidation de la relation.
Phase 5 : Exploitation de la relation de confiance :
On exploite la relation de confiance afin d'obtenir une action frauduleuse de la part de la cible, sans que celle-ci le réalise.
Phase 6 : Debrief
On ramène la cible à un état émotionnel positif / neutre dans le but d'éviter toute forme de culpabilisation ou de sentiments négatifs. Ces derniers pourraient entraîner un refus ultérieur ou une alerte.

Selon les vecteurs d'attaques choisis, les phases 4 et 5 peuvent être imbriquées et la phase numéro 6 peut ne pas avoir lieu. Prenons un exemple de scénario fictif.

Nous sommes les attaquants et nous voulons des informations sur l'administrateur système d'une TPE spécialisée dans l'innovation médicale afin de s'introduire dans le système d'information pour y dérober des informations confidentielles.

Lors de la surveillance et des recherches préliminaires, nous avons remarqué que l'un des employés du service logistique était un grand supporter de l'équipe de rugby de sa ville. Il assistait à tous les matchs de son équipe favorite.

Il n'est pas difficile de se rendre à un match en feignant être fraîchement arrivé en ville et vouloir s'intégrer et rencontrer les autres habitants à travers leur passion commune pour ce sport.

De discussion en discussion et en orientant un peu cette dernière lors des troisièmes mi-temps, nous récoltons de plus en plus d'informations sur l'administrateur système, adresse, horaires de travail et autres habitudes (il n'est pas trop du matin par exemple), prénom de ses quatre enfants (dont un avec un prénom rare et compliqué), goûts musicaux ou culinaires, etc.

Ainsi, il nous est facile de créer un mail de *spear phishing* sur mesure pour créer un accès au réseau de l'entreprise, mais surtout découvrir le mot de passe administrateur, qui n'était autre que le prénom rare de son enfant, de l'Active Directory, et se créer un accès discret au lieu de stockage des documents sensibles.

Par la suite, nous avons gardé le contact avec le fan de rugby jusqu'à ne plus avoir besoin de lui et disparaître dans la nature en prétextant un énième déménagement professionnel, afin de ne pas éveiller les soupçons chez lui.

La mission est réussie. Nous avons récupéré les informations nécessaires pour notre espionnage industriel, et nous n'avons pas été repérés par notre victime.

Dans cette courte fiction, nous retrouvons bien les six phases de notre attaque :
*Phase 1 :*
Le but de l'attaque : dérober des informations confidentielles.
*Phase 2 :*
Recherche d'information :
Observation des utilisateurs, recherche sur une cible potentielle,
*Phase 3 :*
Comment va-t-on avoir accès au réseau de la TPE ? Comment récolter des informations auprès du logisticien ?
*Phase 4 :*
Prise de contact et discussion lors des matches et des troisièmes mi-temps.
*Phase 5 :*
Création d'un *spear phishing*, récupération des données confidentielles.
*Phase 6 :*
Rompre le contact de manière délicate avec le logisticien. Le but : ne pas éveiller les soupçons.

L'utilisation de l'IA peut se retrouver à chacune de ces phases (Mouton & al. 2014 [16]), facilitant l'élicitation de données sur le profil des cibles, amplifiant les effets de l'attaque et surtout facilitant l'accès aux attaques par *social engineering* à beaucoup plus de personnes.

**4. Contre-mesure existante et contrepartie / efficacité**

La méta-analyse de Syafitri et al, 2022 [3], concernant la prévention du *social engineering*, semble montrer des résultats.

On constate que c'est la nouveauté de l'attaque qui crée son opportunité, et conditionne son succès : une fois largement informée, la population sait réagir et éviter la chausse-trappe, et il est alors nécessaire de renouveler la stratégie afin de surprendre à nouveau. L'apprentissage est donc un élément de la solution, donc la formation et la prévention des axes à privilégier.

La prévention contre le *social engineering* peut dès lors revêtir de multiples formes : *serious games*, entraînement par le test (faux mail de *phishing* par exemple), formations dédiées.
Devant une telle diversité, il peut être utile de connaître les performances et les impacts des différentes méthodes de prévention : niveaux d'efficacité, contextes

d'utilisation, limites, afin de cibler les bons outils au bon moment dans le bon contexte.
On remarque que les actions de préventions menées paraissent principalement axées sur la sensibilisation, parfois accompagnées d'entraînements, sur des vecteurs d'attaques spécifiques (exemple : *phishing, vishing*, de plus en plus fréquemment en lien avec les évolutions multimodales de l'IA).
D'après (Mouton et al. 2015 [4]) nous avons également des processus d'aide à la détection tels que le SEADMv2 (Security Education Training Awareness) ou des propositions alliant les capacités de détection des utilisateurs seniors avec du ML (Burda 2020 [5]).

Les actions de prévention, comme la sensibilisation ou l'entraînement sur des vecteurs d'attaques spécifiques, semblent avoir des effets à long terme discutables. En effet, M. Junger et al, 2017 [9] montrent que l'augmentation des compétences de défense contre le *social engineering* n'est pas automatiquement corrélée à chaque nouvelle séance de formation.

Une étude d'Olivier de Casanove et Florence Sèdes 2022 [7] portant sur l'amélioration des programmes SETA (Security Education Training Awareness) a déjà proposé une méthodologie d'amélioration de la prévention. Elle s'appuie notamment sur l'utilisation de l'outil PDCA (Plan – Do – Check – Act) afin de créer un framework d'amélioration continue à l'aide de collecte et d'analyse des données lors des différentes campagnes de prévention et de sensibilisation.

La méta-analyse de Bullee et Junger 2020 [20] et l'étude de P. Kumaraguru [19] apportent des pistes de travail dans le cadre de la préparation d'un plan de prévention. Il faut cependant noter que les interventions sont principalement exécutées en laboratoire et sur un seul vecteur d'attaque à la fois. Néanmoins, elles respectent dans l'ensemble les principes de prévention et de formation validés par les spécialistes en science de l'éducation, aussi bien pour les adultes [32] que pour les enfants [33], ouvrant ainsi la voie à une possibilité de prévention dès le plus jeune âge.

Nous pouvons essayer d'en déduire les recommandations suivantes :

➔ Former tout le monde, pas seulement celles et ceux ayant échoué aux tests.
➔ Tester avec un feedback immédiat pour expliquer les erreurs commises.
➔ Travailler sur des sujets spécifiques.
➔ Préférer une modalité d'intervention dynamique avec composante verbale.
➔ Gamifier pour faciliter l'apprentissage.
➔ Avoir recours à des illustrations plutôt qu'une simple présentation ou du texte seul.
➔ Assimiler le mécanisme des URL qui a un effet relativement important.
➔ Eviter les messages d'avertissement, quasi-inutiles.
➔ Sensibiliser, entraîner et tester régulièrement.

Une recommandation globale et partagée consiste à prôner une approche du *social engineering* inclusive [28].

**5. Défis et orientations futures**

Comme évoqué par Schmitt et al. [17] et Falade et al. [18], la lutte contre le *social engineering* amplifié par IA et ML ne pourra se passer de la prévention humaine. Elle nécessitera également la collaboration des différents secteurs afin de développer des outils dont des IA pour venir soutenir la défense, par exemple en permettant de découvrir les variations, même très légères, laissées par les IA génératives dans leurs productions.

Malgré les investissements pour établir des stratégies efficaces contre les attaques, les méthodes de détection existantes sont limitées et les contre-mesures inefficaces pour faire face au nombre croissant d'attaques, en raison des vulnérabilités technologiques et humaines exploitées. Parce que l'humain est un défi pour la sécurité de tout réseau, il est important de développer les programmes de formation pour les employés et au-delà tout citoyen, cible potentielle, en investissant dans l'éducation à la cybersécurité :
- Quelles sont les meilleures pratiques pour sensibiliser et former les utilisateurs aux risques du *social engineering* ?
- Comment les organisations peuvent-elles collaborer pour lutter contre la cybercriminalité ?
- Quelles sont les implications éthiques de l'utilisation de l'IA dans la cybersécurité ?
- Comment garantir la protection de la vie privée et de la confiance numérique face aux menaces du *social engineering* ?

Les politiques de déploiement s'avérant insuffisantes, la nécessité de réfléchir à des étapes amont se fait jour : savoir anticiper, imaginer en se mettant « à la place » des attaquants, détecter précocement signaux faibles et *outliers* et réagir promptement face à la délinquance informatique sont alors des questions prioritaires nécessitant une approche axée sur la prévention et la coopération.

## 6. Conclusion

Dans cette synthèse, nous avons donné un aperçu des attaques de *social engineering*, des techniques de détection existantes, et des contre-mesures, à l'efficacité limitée, pour des raisons technologiques ou humaines : un système de sécurité robuste peut être facilement contourné par une simple attaque de *social engineering*. De telles attaques ont augmenté en intensité et en nombre et causent d'importants dommages émotionnels et financiers, depuis les institutions publiques jusqu'au particulier, en passant par les entreprises de toutes tailles.

Cette revue de littérature a été synthétisée pour donner un aperçu des différentes formes d'attaques existantes ainsi que des leviers humains sur lesquels elles s'appuient.

Le *social engineering*, augmenté des technologies d'IA qui se généralisent (multimodalité, vidéo, audio…) reste une menace importante en constante évolution, qui ouvre de nouvelles perspectives : peut-on réellement former efficacement sur tous les vecteurs d'attaques connues ou non ? que se passe-t-il lorsqu'un utilisateur ou une utilisatrice se retrouve face à une méthode d'attaque jusqu'alors inconnue ?

La prévention et la formation continue des utilisateurs et des utilisatrices reste un socle indispensable pour une défense efficace, en synergie avec les moyens de défense technique. Le *social engineering* est un défi majeur pour la sécurité des SI et la cybersécurité. Une approche multidimensionnelle et collaborative est nécessaire pour contrer cette menace en constante évolution, la sensibilisation, la formation et la recherche de solutions innovantes s'avérant essentielles pour protéger les individus, les organisations et les sociétés.

Des approches pluridisciplinaires seront nécessaires pour améliorer les programmes de prévention, développer de nouveaux outils et essayer de prendre de l'avance sur les cybercriminels. L'impact du *social engineering* sur la vie privée et la confiance numérique est essentiel à prendre en compte, de même que les implications éthiques de l'utilisation de l'IA dans la cybersécurité.

## 7. Références

[1] Verizon : 2023 Data Breach Investigations Report
https://www.verizon.com/business/resources/reports/dbir/2023/summary-of-findings/
[2] Wang et al., 2021, Social Engineering in Cybersecurity: Effect Mechanisms, Human Vulnerabilities and Attack Methods, IEEE Access, 2021, vol. 9, pp. 11895-11910
DOI. 10.1109/ACCESS.2021.3051633
[3] Syafitri et al, 2022, Social Engineering Attacks Prevention: A Systematic Literature Review,IEEE Access, 2022, Vol. 10, p. 39325 - 39343
DOI.10.1109/ACCESS.2022.3162594

[4] Mouton et al. 2015, Social Engineering Attack Detection Model: SEADMv2, 2015 International Conference on Cyberworlds (CW), 2015, p. 216 - 223 DOI 10.1109/CW.2015.52
[5] Burda 2020, Don't Forget the Human: a Crowdsourced Approach to Automate Response and Containment Against Spear Phishing Attacks, 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), 2020, p. 471 - 476 DOI 10.1109/EuroSPW51379.2020.00069
[6] Perrotte et Croutte, 2021, L'escroquerie en ligne et à la téléphonie en France : ampleur du phénomène et profils des victimes, CREDOC, 2021
[7] Olivier de Casanove, Florence Sèdes, 2022, Applying PDCA to Security, Education, Training and Awareness Programs. HAISA, 2022, p.39 - 48 DOI : hal-03759487
[9] M. Junger et al., 2017, Priming and warnings are not effective to prevent social engineering attacks, Computers in Human Behavior, 2017, vol 66, p. 75-87
[10] Yasin et al. 2019, Contemplating social engineering studies and attack scenarios: A review study, Security and privacy, vol. 2, p.73
[11] Party et Rajendran, 2019, Identification and prevention of social engineering attacks on an enterprise, 2019 International Carnahan Conference on Security Technology (ICCST), IEEE, p. 1-5
[12] S. Kaushalya, R. Randeniya, and A. Liyanage, 2018, "An overview of social engineering in the context of information security," in 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2018: IEEE, pp. 1-6
[13] Wang et al., 2020, Defining Social Engineering in Cybersecurity, IEEE Access, vol. 8, p. 85094 - 85115
[14] "Traité de psychologie sociale", Laurent bègue, Olivier Desrichard, Deboeck superieur
[15]https://www.enderi.fr/Tupolev-Tu-144-l-espionnage-industriel-au-coeur-de-la-guerre-froide_a471.html
[16] Mouton et al. 2014, Social Engineering Attack Framework, 2014 Information Security for South Africa (ISSA), IEEE, P. 1 - 9 - DOI 10.1109/ISSA.2014.6950510
[17] Schmitt et Flechais, 2023,Digital Deception Generative Artificial Intelligence in Social Engineering and Phishing. SSRN Electronic Journal, DOI https://doi.org/10.48550/arXiv.2310.13715
[18] Falade, 2023, Decoding the Threat Landscape ChatGPT, FraudGPT, and WormGPT in Social Engineering Attacks, Intl Journal of Scientific Research in Computer Science, Engineering and Information Technology, p. 185 - 198
[19] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong. 2010 "Teaching Johnny Not to Fall for Phish", in Proc. of the ACM Transactions on Internet Technology, 10(2), pp. 1-31
[20] Buller et Junger, 2020, How effective are social engineering interventions? A meta-analysis. Information & Computer Security- DOI 10.1108/ICS-07-2019-0078
[21] J. Kävrestad, A. Hagberg, M. Nohlberg, J. Rambusch, R. Roos and S. Furnell, 2022, "Evaluation of Contextual and Game-Based Training for Phishing Detection," Future Internet, vol. 14, p. 104,

[22] J. M. Hatfield, 2018, ‘‘Social engineering in cybersecurity: The evolution of a concept,’’ Comput. Secur., vol. 73, pp. 102–113,
[23] Slade, John A. 1929. “Law and Psychology.” The Journal of Abnormal and Social Psychology 24(2): 212-216.
[24] S. Lohani, 2019, "Social engineering: Hacking into humans," International Journal of Advanced Studies of Scientific Research, vol. 4, no. 1.
[25] J.-H. Cho, H. Cam, and A. Oltramari, 2016, "Effect of personality traits on trust and risk to phishing vulnerability: Modeling and analysis," in 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2016: IEEE, pp. 7-13
[26] F. Sèdes, 2024, “Les femmes dans la cybersécurité”. https://www.amue.fr/publications/la-collection-numerique
N° 31 Sécurité des SI : La cybersécurité au cœur de la stratégie de l'ESRI 02/2024
[27] A. Kumar, M. Chaudhary, and N. Kumar, 2015, "Social engineering threats and awareness: a survey," European Journal of Advances in Engineering and Technology, vol. 2, no. 11, pp. 15-19.
[28] https://infosecawareness.in/concept/social-engineering
[29]Joseph M. Hatfield, 2023, There Is No Such Thing as Open Source Intelligence, Intl Journal of Intelligence and CounterIntelligence Vol. 37, 2024, p. 397-418.
[30] Fatima Salahdine, Naima Kaabouc, 2019. Social Engineering Attacks: A Survey. Future Internet 2019, 11, 89; doi:10.3390/fi11040089.
[31] Olivier de Casanove, Florence Sèdes, 2022. Malicious Human Behaviour in Information System Security: Contribution to a Threat Model for Event Detection Algorithms. Foundations and Practice of Security - 15th International Symposium, FPS 2022, Ottawa, ON, Canada, December 12-14, 2022: 208-220.
[32] Clark, R.C. and Mayer, R.E. (2016), E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning, John Wiley and Sons
[33] Franck Ramus et al., 2021, MOOC : La psychologie pour les enseignants. 2021. https://www.fun-mooc.fr/fr/contributeurs/franck-ramus/
[34] Influence et Persuasion, Robert Caldini 2014, Pocket 2024
[35] Petit Traité de Manipulation à l’usage des honnêtes gens. Robert Vincent Joule, Jean-Léon Beauvois, PUG 2022

# Analyser et écrire la science multidisciplinaire dans un réseau d'hypertextes sémantiques avec Wicri

**Jacques Ducloy**[1]

*1. Laboratoire Paragraphe, Université Paris 8*
*2 rue de la Liberté - 93526 Saint-Denis cedex, France*
*Jacques.Ducloy@univ-lorraine.fr*

*RESUME. Le projet WICRI travaille sur une alternative à la galaxie Wikipédia pour les communautés scientifiques. Il prend également en compte des besoins d'analyse de corpus avec une boîte à outil XML pour la création de systèmes d'informations, notamment dans la perspective de développer des bases de données bibliographiques. Enfin, dans ce réseau culturel, les rééditions hypertextes d'ouvrages ancien et de manuscrits sont particulièrement démonstratives des applications avancées dans les humanités numériques (avec notamment une bibliothèque numérique sur la Chanson de Roland).*

*ABSTRACT. The WICRI project aims at a potential alternative to the Wikipedia galaxy for scientific communities. It also takes into account corpus analysis needs with an XML toolbox for the creation of information systems. Finally, in this cultural network, hypertext reissues of old works and manuscripts are particularly suited to advanced applications in digital humanities (including a digital library on the Song of Roland).*

*Mots-clés : Semantic Mediawiki, Ingénierie XML, Réseau de Wikis, Exploration de corpus, Humanités numériques, Chanson de Roland*

*KEYWORDS: Semantic Mediawiki, XML engineering, Wiki network, Corpus exploration, Digital humanities, Chanson de Roland*

## 1. Introduction

Le projet présenté est issu de travaux de R&D menés à l'INIST en 1990. Il s'agissait de construire un système d'IST (Information Scientifique et Technique) compétitif au niveau international. Il concernait notamment les bases Pascal et Francis (500.000 analyses par an par environ 400 ingénieurs) avec des mécanismes d'indexation assistée, intégrant les spécificités de chaque secteur scientifique, et des mécanismes de coopérations prenant en compte l'ensemble des besoins informationnels de la recherche. Mais en 1992, le CNRS a engagé un virage à 180° en visant un groupe commercial. Le département de R&D a été dissous. Les bases Pascal et Francis ont été arrêtées en 2015. Dans la même période, rien que sur la santé, aux USA, la *National Library of Medicine* (NLM) a doublé sa capacité de production (1.000.000 d'analyses par an par 750 ingénieurs, et un réseau de 8.000 collaborateurs). De même, dans les années 60, le CNRS avait lancé le dictionnaire du Trésor de la Langue Française qui a quasiment disparu face à la domination de la Wikimedia Foundation (750 personnes à San Francisco, plus de 200 millions de dollars de chiffre d'affaires).

Nous avons donc décidé d'engager une réflexion allant de l'ingénierie aux pratiques avancées des chercheurs et des partenaires de la recherche. Nous disposons d'un démonstrateur qui apporte un début de preuve de concept. Il permet déjà de réaliser des premiers services opérationnels et de mener de multiples expérimentations, soit technologiques, sur les systèmes de réseau hypertexte, soit éditoriales, comme l'écriture hypertexte collaborative ou multidisciplinaire.

## 2. Histoire des projets Dilib et Wicri

En 1991, le premier résultat obtenu à l'INIST a été une boîte à outil SGML (en préfiguration d'XML). Reprise par le Loria sous l'appellation Dilib, elle permettait de réaliser des serveurs d'exploration de corpus bibliographiques hétérogènes. On y associait des mécanismes de classification à des fonctions plus classiques, de type moteur de recherche. Il était ainsi possible de réaliser des applications de taille modeste avec un grand niveau d'interdisciplinarité (comme par exemple une base iconographie et bibliographique sur l'art nouveau) ou des services à volumétrie conséquente (l'intégralité des bases Pascal et Francis).

Un deuxième axe a été initialisé par une réflexion autour de Wikipédia qui apportait des éléments de réponse à la gestion des flux de contributions rencontrés dans la production des bases. Elle nous a amené à travailler sur une alternative à cette encyclopédie pour la production d'informations produites par la recherche (et donc souvent nouvelles et non sourcées). Nous avons envisagé une collection d'encyclopédies thématiques qui pourraient être pilotées et modérées par des comités scientifiques. En 2008, un démonstrateur (Wicri) a donc été construit sous la forme d'un réseau de wikis dopés par des mécanismes d'annotation sémantique

(avec Semantic MediaWiki). Grâce un financement CPER le réseau a acquis une dimension multidisciplinaire, notamment dans les sciences liées à la santé et l'environnement.

Dans le cadre d'ISTEX, le projet LorExplor en 2013 a permis de rapprocher les deux approches. Plus précisément, des serveurs d'exploration ont été intégrés à la base Wiki sémantique. En amont, des mécanismes de curation sont basés sur des formalisations gérées dans les wikis. En aval, la bibliothèque XML offre des procédures de génération de modèles en utilisant par exemple les outils de visualisation géographique de Wikipédia. La bibliothèque XML permet également de développer des robots pour assister les actions éditoriales.

La fin des financements ISTEX a réduit les capacités de coopération. Nous avons donc recherché des thématiques que nous pouvions explorer sans l'obligation technique de recourir à une expertise extérieure (comme la santé). Une première série d'expériences en musique a amené à enrichir notre panoplie de services avec des rééditions hypertexte d'ouvrages avec des éléments musicaux (comme le Dictionnaire de Jean-Jacques Rousseau). En 2020, une nouvelle étape a été franchie avec une bibliothèque numérique sur la Chanson de Roland. Ici, pratiquement chaque document (strophe d'un manuscrit, chapitre d'une édition critique, article de recherche, partition) demande un traitement numérique spécifique. Chaque mot d'un manuscrit (ou d'une note de philologue) peut devenir un élément hypertexte dont les explications peuvent de développer dans plusieurs wikis.

### 3. Le démonstrateur Wicri

Le réseau actuel est un ensemble encyclopédique étendu (avec par exemple des rééditions d'ouvrages et des extraits de bases de données) développé sur 150 wikis. Il offre également des applications stabilisées comme la revue les mots de l'agronomie de l'INRAE.

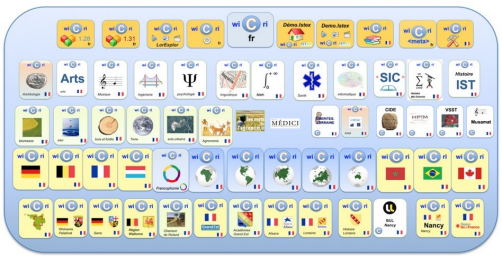



*Figure 1. Le réseau Wicri*

En termes de volumétrie, ce réseau expérimental contient 200.000 pages wiki (avec 40.000 articles conséquents et 13.000 fichiers multimédia). En complément, 150 serveurs d'exploration donnent accès à plus d'un demi-million de documents.

## 4. Démonstrations proposées

Ce réseau propose un très vaste champ de démonstrations. De façon générale il est possible de voir la multiplicité des relations sémantiques et les mécanismes de cohérence dans le réseau de wikis. Au-delà de cette base, quatre applications significatives sont proposées. Sur le COVID une vingtaine de serveurs d'explorations ont été réalisés avec une procédure rapide de mise en place (quelques minutes). Sur la Chanson de Roland nous avons construit un ensemble qui contient déjà près de 4.000 pages significatives (chapitres d'ouvrage, versets de manuscrit, analyses critiques, etc.). L'Histoire de l'Information scientifique et technique est une thématique en cours de démarrage et qui doit se décliner dans l'ensemble des wikis du réseau. Enfin, un travail sur l'intégration de la très vaste Histoire naturelle de Buffon dans une bibliothèque hypertexte vient d'être initialisé.

## 5. Remarques et perspectives

Le démonstrateur Wicri montre la faisabilité d'un déploiement de l'IST française pour retrouver une dimension internationale. Il montre aussi l'intérêt de cette technologie dans les applications de la recherche où les approches classiques des systèmes d'information échouent ou donnent lieu à des développements particulièrement laborieux.

Nous démarrons une nouvelle étape dans laquelle nous allons étudier une répartition du réseau Wicri sur plusieurs sites physiques. Avec des moyens limités (un seul permanent, le retraité auteur de cet article), nous allons démarrer avec 2 sites dans un premier temps et avec une augmentation du nombre des wikis (pour prendre en compte, par exemple, l'ensemble des régions françaises).

Pour aborder une vraie dimension internationale, compétitive avec la volumétrie de la Wikimedia Fundation, il faudrait passer à quelques milliers de wikis sur une centaine de sites. Cet objectif nous semble difficile et ambitieux mais techniquement abordable pour la communauté universitaire. Mais la base actuelle offre déjà une infrastructurel de formation, expérimentation et même de services…

## 6. Liens et bibliographie

Liens vers le réseau Wicri :

– le wiki d'accueil : https://wicri-demo.istex.fr/Wicri/Wicri/fr ;

– Une reproduction hypertexte de cet article avec accès aux démonstrations.
https://wicri-demo.istex.fr/Wicri/Sic/fr/index.php/INFORSID_Nancy_(2004)_Ducloy

Bibliographie du projet Wicri :

https://wicri-demo.istex.fr/Wicri/Wicri/fr/index.php/Bibliographie_du_projet_Wicri

# Comment gérer les risques liés à l'interconnexion des systèmes tiers dans un système de traitement de données configuré par graphe ?

**Jean-Sébastien Fest , Anthony Bonnemaire , Jean Bort , Philippe Garnier**

*Emvista*
*10, rue Louis Breguet*
*34830 Jacou, France*
*prenom.nom@emvista.com*

RÉSUMÉ. *Les systèmes de traitement de données sont souvent conçus pour répondre à un besoin ou une famille de besoins exprimés par un métier. Une des limites de tels systèmes est leur capacité à s'adapter d'une part à l'évolution du besoin et d'autre part aux risques techniques omniprésents tels qu'une erreur retournée par un système tiers nécessaires au traitement. Cet article décrit les démonstrations qui seront réalisées lors de la conférence pour montrer comment ces limites ont pu être levées grâce à Bravo, notre framework low code qui a pour ambition d'être le plus générique possible pour le développement d'application.*

ABSTRACT. *Data processing systems are often designed to respond to a need or a family of needs expressed by a business. One of the limits of such systems is their ability to adapt on the one hand to changing needs and on the other hand to omnipresent technical risks such as an error returned by one of the modules necessary for processing. This article describes the demonstrations we will do during the conference in order to show how these limits could be lifted thanks to Bravo, a low code framework which aims to be as generic as possible.*

MOTS-CLÉS : *boîte à outils, traitement de données, risk management*

KEYWORDS: *framework, data processing, gestion du risque*

## 1. Introduction et principes de base

La conception de boîtes à outils (ou *framework*) permettant de créer des applications fondées sur des systèmes tiers, tels que Akka (Roestenburg *et al*., 2016), Apache NiFi [1], n8n [2] ou Flink [3] constitue un enjeu crucial à l'heure où la quantité de modèles d'intelligence artificielle (IA) augmente considérablement alors que leur interopérabilité n'est pas assurée. De façon générale, l'interconnexion des systèmes (IA et autres) est rarement assurée au sein des processus métiers. Dans ce cadre, nous avons conçu et développé Bravo, un framework *low code* original qui a pour ambition d'être le plus générique possible pour le développement d'applications, en particulier celles fondées sur des systèmes tiers et pour lesquelles il est important de concevoir la gestion des risques liés à l'interconnexion desdits systèmes.

L'organisation internationale de normalisation [4] (ISO) définit un **processus** comme étant un ensemble d'activités corrélées ou interactives qui transforment des éléments d'entrée en éléments de sortie. Par ailleurs, une **procédure** est la manière spécifiée d'effectuer un processus. Alors que la plupart des systèmes de traitement de données impose des contraintes relatives à une procédure, Bravo conserve la distinction entre la procédure et le processus dans sa modélisation (cf. Fig. 2). C'est un point fondamental qui augmente la généricité recherchée par Bravo et qui est un socle à la gestion des risques. D'une part, la procédure est modélisée par un graphe orienté (nommé dans la suite "graphe de procédure"). Les nœuds de ce graphe représentent des processus tels que la compilation ou la sauvegarde. La caractéristique fondamentale de ces nœuds est qu'ils sont des points de persistance du traitement (ils ne représentent pas le traitement en soi). Les arcs indiquent l'ordre de traitement des nœuds. D'autre part, les processus impliqués dans la procédure sont également modélisés sous forme de graphes (nommés dans la suite "graphes de processus"). Les nœuds de ces graphes sont des fonctions telles que la lecture et l'écriture dans une base de données et les arcs indiquent l'ordre de traitement des nœuds. Enfin, pour modéliser le flux, le graphe de procédure connecte les graphes de processus entre eux grâce à des nœuds nommés "connecteurs" qui permettent de spécifier une itération ou une fusion de résultats, entre autres choses, à l'issue d'un graphe de processus.

À travers le graphe de procédure et ses graphes de processus, un objet unique circule. Cet objet permet de définir la tâche à réaliser et contient toutes les informations pour exécuter le graphe de procédure. Il contient les données sources à traiter, les traces des activités (i.e. *logs*) ainsi que les résultats à l'issue de chaque étape. Chaque nœud fait évoluer le statut de cet objet ainsi que son réservoir de données et son journal.

1. https://nifi.apache.org/
2. https://n8n.io/
3. https://flink.apache.org/
4. https://www.iso.org/fr/home.html

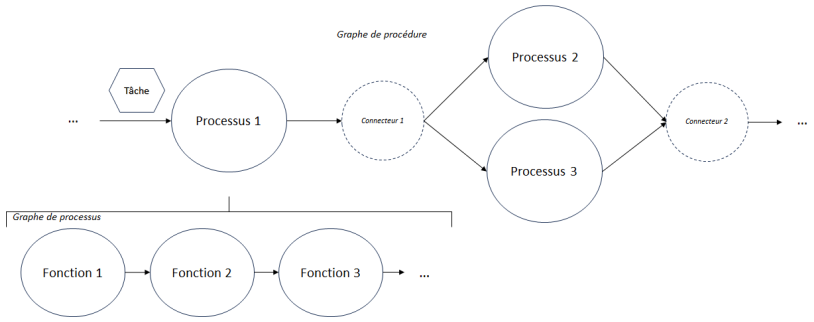



FIGURE 2. *Représentation d'un graphe de procédure et ses graphes de processus*
FIGURE 1.

## 2. Gestion des risques liés à l'interconnexion avec des systèmes tiers

Développer des applications fortement connectées avec d'autres systèmes et dont les traitements sont dépendants de l'orchestration de processus asynchrones non garantis est complexe. Cela fait partie des nouveaux challenges du développement logiciel (Charette, 2017) (Kleppmann, 2019). Nous évoquerons lors de la session "démonstration" comment Bravo gère la pression (cf. section 2.1) et les échecs (cf. section 2.2) des systèmes tiers.

### *2.1. Gestion de la pression sur les systèmes tiers*

Une attention particulière doit être portée aux requêtes (de lecture ou d'insertion) à des systèmes tiers, qu'ils soient des services internes ou partenaires afin de ne pas les surcharger. Bravo dispose d'un régulateur qui permet de spécifier le nombre maximum de graphes de procédures. Bravo permet également de définir par configuration le nombre maximum d'appels concurrents exécutés.

### *2.2. Gestion des échecs de communication avec les systèmes tiers*

L'intégration d'appels à des systèmes tiers pour lesquels ni la performance ni l'accessibilité n'est maîtrisée nécessite la mise en place de stratégies également configurables. Par exemple :

– dans le cas où un dysfonctionnement apparaîtrait suite à l'appel à un système tiers, l'utilisateur peut configurer s'il souhaite que l'appel soit relancé ou si les résultats de l'appel défectueux doivent être utilisés ;

– l'utilisateur peut décider si un échec est bloquant ou pas lors de l'exécution du graphe de procédure.

Lors de l'appel en échec vers un système tiers, Bravo implémente plusieurs stratégies de réessai. Le système va réessayer plusieurs fois d'appeler le système tiers avec une stratégie d'incrément temporel (par exemple de l'ordre de quelques secondes) entre chaque essai. En cas d'échec des essais, la tâche est mise dans une pile de tâches n'ayant pas abouti et sera relancée selon la configuration des relances. Il est par exemple possible de configurer un nombre maximum de relances qui, une fois atteint, déclenche des alertes vers des sondes de surveillance et par e-mail.

## 3. Démonstrations

Dans le cadre de la démonstration, nous montrerons à travers deux cas d'usage comment Bravo gère les risques liés à l'interconnexion avec des systèmes tiers. Le premier cas d'usage témoigne de l'efficacité de la gestion de risques proposée dans le cadre de l'analyse de tableurs. Le second cas d'usage témoigne d'une procédure plus complexe qui analyse les e-mails d'une boîte de réception au fur et à mesure de leur réception. Lors de la démonstration, nous simulerons des échecs au sein de la procédure et montrerons comment ils sont gérés automatiquement par le système.

– **Cas d'usage "analyse de verbatims".** Il s'agit pour cette démonstration d'analyser des milliers de lignes contenant des données numériques et textuelles en utilisant Bravo. Plusieurs services d'analyse de contenu, y compris des modèles d'IA, sont utilisés pour aboutir au résultat de l'analyse : un tableau de bord permettant de visualiser les termes clés, les concepts, les émotions et les opinions véhiculés dans les cellules du tableur. Nous montrerons comment le système empile les demandes de façon sécurisée (sans perte de requêtes), comment il respecte les contraintes du nombre d'appels possible aux systèmes tiers, et comment il peut reprendre les traitements en échec lorsqu'un appel n'a pas abouti sans relancer les traitements qui ont abouti.

– **Cas d'usage "analyse d'e-mails".** Dans notre application d'analyse d'e-mails, un utilisateur connecte un compte e-mail à l'application développée avec Bravo. L'utilisateur configure le graphe de procédure avec une interface de configuration afin que ses e-mails soient automatiquement lus puis analysés par un service d'analyse sémantique dans le but de classer les e-mails. Nous montrerons comment réagit le système face aux échecs simulés telles que la communication avec le serveur d'authentification de l'utilisateur (authentification déléguée), la communication avec le serveur e-mail (par exemple Gmail ou Microsoft Graph), la communication avec les services Web d'analyses et la communication avec la base de données OpenSearch.

## Bibliographie

Charette R. N. (2017). It's fatal amnesia. *Computer*, vol. 50, n° 02, p. 86–91.

Kleppmann M. (2019). *Designing data-intensive applications.* English.

Roestenburg R., Williams R., Bakker R. (2016). *Akka in action*. Simon and Schuster.

# FLOC: outil de mesure énergétique multi-composants

**Hernan Humberto ALVAREZ VALERA** [1], **Franck RAVAT** [2], **Jiefu SONG** [2], **Philippe ROOSE** [4], **Nathalie VALLES-PARLANGEAU** [4]

1. *Domolandes, Saint Geours de maremne, France*
   *humberto.valera@domolandes.fr*
2. *IRIT (CNRS 5505)- Université Toulouse Capitole, Toulouse, France*
   *prenom.nom@irit.fr*
3. *LIUPPA,Université de Pau et des Pays de l'Adour, Anglet, France*
   *prenom.nom@univ-pau.fr*

*RÉSUMÉ.*
*Mesurer la consommation énergétique des applications est un préalable nécessaire à toute stratégie de réduction de l'empreinte écologique de l'IT. Cette consommation est directement liée à la charge générée sur différents composants matériels par des applications.*
*Cet article présente FLOC, un outil conçu pour mesurer et fournir des profils complets de consommation d'énergie des applications et des ensemples d'applications qui conforment un système complexe. FLOC couvre des composants matériels clés tels que le CPU, la RAM, les cartes réseau et les dispositifs de stockage.*
*Nous avons démontré la pertinence de FLOC sur un benchmark de Data Lake.*

*MOTS-CLÉS : consommation énergétique globale, outils de mesure logiciels, systèmes big data*

## 1. Introduction

Pour atteindre zéro émission d'ici 2050, une réduction de 55% de la consommation électrique est attendue. Concernant le numérique, l'un des leviers est lié à l'exécution des logiciels, dont la consommation énergétique augmente de 7% chaque année [1]. Afin de réduire efficacement cette consommation, il s'agit d'abord de la mesurer de manière pertinente. C'est-à-dire, il faut considérer les composants matériels impliqués

1. https://www.digitalinformationworld.com/2020/02/the-global-energy-consumption-of-information-technologies-infographic.html

dans l'exécution d'une application unique ou d'un ensemble d'applications formant un système complexe, tout en incluant leurs processus, sous-processus et threads.

Par exemple, concernant les Big Data Analytics, et dans le cas qui nous concerne, les Lacs de Données, les traitements dans les différentes zones impactent différemment les composants matériels. Durant la phase d'ingestion, la carte réseau gère le transfert des données, tandis que les dispositifs de stockage garantissent leur pérennité et accessibilité. Lors de la phase de stockage, ces dispositifs prennent en charge l'organisation, l'indexation, et la récupération des données. Enfin, durant la phase de nettoyage des données, le CPU est intensivement utilisé, tandis que la RAM offre un stockage temporaire pour les données en cours de traitement et les résultats intermédiaires. Ainsi, pour élaborer des méthodes d'optimisation holistiques, comment pouvons-nous mesurer l'intégralité de ce processus complexe ? Comment déterminer quelle phase est la plus coûteuse en terme énergétiques ? Quel composant matériel est le plus gourmand en énergie?

Actuellement, deux méthodes de mesure existent : d'une part, les capteurs physiques évaluent précisément la consommation des dispositifs (ordinateurs, serveurs, etc.), mais ne distinguent pas la consommation des applications spécifiques ou celle des composants matériels distincts (CPU, RAM, etc.) (Jay *et al*., 2023). D'autre part, des outils logiciels permettent d'évaluer la consommation d'un ensemble d'applications. Certains sont développés comme une couche au-dessus des interfaces de bas niveau des processeurs pour évaluer la consommation des processus (Noureddine, 2022 ; Petit, 2023) ou des fragments du code source (Hähnel *et al*., 2012). Ces derniers ciblent seulement le CPU, négligeant des composants clés comme la RAM ou les périphériques d'entrée/sortie pour une analyse complète de la consommation énergétique (Alvarez-Valera *et al*., 2024) . D'autres approches considèrent divers composants matériels mais se limitent à l'analyse d'une charge de travail spécifique pendant une période déterminée et, à notre connaissance, n'ont pas maintenu la compatibilité avec les nouvelles versions des systèmes d'exploitation. Enfin, certaines méthodes étudient la consommation de l'intégralité du trajet d'une requête client dans des environnements distribués (Anand *et al*., 2023), mais elles ne sont pas adaptées pour analyser le comportement énergétique d'applications ou de systèmes complexes, tels que les systèmes de big data analytics, sur le long terme.

Nous proposons FLOC, un outil pour mesurer de façon complète la consommation énergétique d'applications individuelles ou des systèmes complexes, prenant en compte leurs processus, sous-processus, et threads, ainsi que divers composants matériels clés tels que le CPU, la RAM, la carte réseau, ou les dispositifs de stockage.

## 2. Fonctionnement de FLOC

FLOC repose sur la mesure de la charge induite des composants matériels lors de l'exécution d'une application, de frameworks ou d'un ensemble d'applications, ainsi que les processus, sous-processus et threads associés. FLOC se base sur des calculs spécifiques à chaque composant pour convertir ces charges en valeurs de puissance en

watts. Ce processus est répété à des intervalles définis $i$ et sur une durée totale $t$ spécifiée. Ainsi, FLOC calcule la puissance moyenne consommée par chaque composant pendant cette période, déterminant ainsi la consommation énergétique totale de l'application en joules (la puissance en watts multipliée par le temps en secondes). Pour le CPU, les calculs se basent sur le temps CPU alloué aux applications. Pour la RAM, ils considèrent le coût énergétique des opérations de lecture et d'écriture, alors que pour les disques et le réseau ce sont les opérations d'entrée/sortie qui sont évaluées.

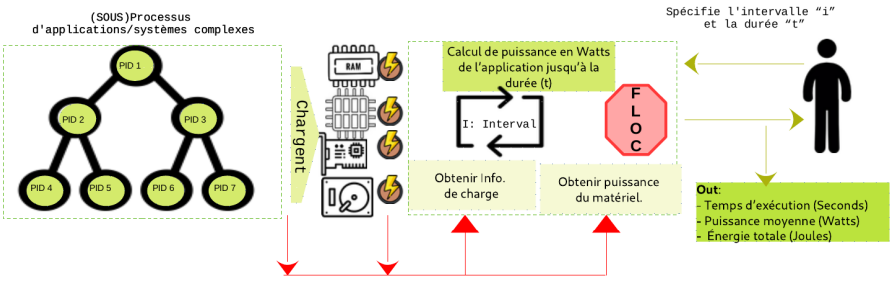



FIGURE 1. *Fonctionnement de FLOC*

Du point de vue de l'utilisateur (Figure 1), il saisit le nom de l'application, le PID du processus ou le nom du framework à évaluer, ainsi que l'intervalle $i$, la durée $t$ et les composants d'intérêt. L'application enregistre la puissance moyenne en Watts (W) et l'énergie consommée en Joules (J) pour chacun des composants spécifiés. Si $t$ est négatif, FLOC fonctionne indéfiniment jusqu'à la fin de tous les PIDs concernants.

## 3. FLOC en action sur un système de gestion et d'analyse de données

Nous avons choisi d'évaluer l'utilité et de demontrer la pertinence de FLOC au travers de la mesure de la consommation d'énergie d'un benchmark spécialement conçu pour évaluer les opérations Big Data dans un lac de données (Sawadogo, Darmont, 2023). C'est un système complexe dont l'exécution génère une charge sur les 4 composants matériels les plus importants, que FLOC est capable d'évaluer. Pendant sa phase d'ingestion, ce benchmark transforme de grands ensembles de données brutes (PDFs d'articles scientifiques et tables SQL) en textes indexables dans Elasticsearch. Il génère ensuite des métadonnées stockées dans Elasticsearch et MongoDB. Il optimise ces données en utilisant Neo4j pour les relations et MongoDB pour le traitement du texte, et applique des analyses de similarité avec SQLite et Neo4j.

L'expérience a consisté à exécuter et évaluer la consommation énergétique de la phase d'ingestion du benchmark, soit 18,75 Go. Cette phase comprend aussi bien l'intégration de données brutes que les mises à jour des métadonnées associées.

La figure 2 présente la consommation énergétique des opérations d'ingestion du benchmark. Elle révèle que, pour l'ingestion de données, le dispositif de stockage est le composant ayant le plus consommé d'énergie avec 197.46 J, soit 39.8% du total, suivi par la RAM à 111.95 J (22.6%) et la carte réseau à 86.39 J (17.4%). Ce qui est

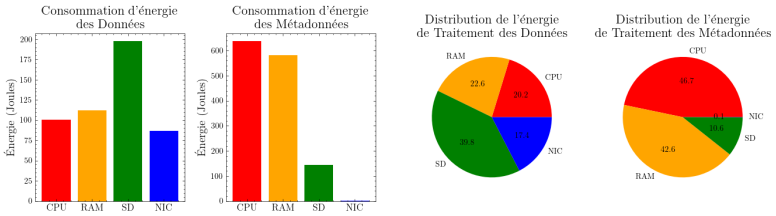



FIGURE 2. *FLOC : Résultats du traitement des données et des métadonnées*

prévisible étant donné que les opérations sont principalement de type E/S. Concernant le traitement des métadonnées, le CPU avec 637.27 J (46.7%) et la RAM avec 581.23 J (42.6%) sont nettement plus énergivores en raison de la complexité des opérations impliquées.

Grâce à **FLOC**, nous spécifions la consommation individuelle des composants matériels pour chacune des opérations effectuées par une application ou un système complexe. Ceci permet d'envisager des stratégies potentielles éco-responsables de répartition de charge, d'optimisation de codage, de traitement de données, etc.

## 4. Travaux futurs

Actuellement, nous améliorons FLOC pour (i) permettre d'intégrer d'autres composants matériels et méthodes de mesure, comme RAPL pour le CPU et (ii) garantir sa compatibilité avec les différentes distributions GNU/Linux. Pour en savoir plus, vous pouvez visiter https://github.com/hhumbertoAv/FLOC/tree/main.

## Bibliographie

Alvarez-Valera H. H., Maurice A., Ravat F., Song J., Roose P., Valles-Parlangeau N. (2024). Energy measurement system for data lake: An initial approach. In *16th asian conference on intelligent information and database systems.* (Accepted for publication)

Anand V., Xie Z., Stolet M., De Viti R., Davidson T., Karimipour R. *et al.* (2023). The odd one out: Energy is not like other metrics. , vol. 3.

Hähnel M., Döbel B., Völp M., Härtig H. (2012). Measuring energy consumption for short code paths using rapl. *SIGMETRICS Perform. Eval. Rev.*.

Jay M., Ostapenco V., Lefevre L., Trystram D., Orgerie A.-C., Fichel B. (2023). An experimental comparison of software-based power meters: focus on cpu and gpu. In *2023 ieee/acm 23rd international symposium on cluster, cloud and internet computing.*

Noureddine A. (2022). Powerjoular and joularjx: Multi-platform software power monitoring tools. In *2022 18th international conference on intelligent environments (ie).*

Petit B. (2023). *scaphandre.* Consulté sur https://github.com/hubblo-org/scaphandre

Sawadogo P. N., Darmont J. (2023). Dlbench+: A benchmark for quantitative and qualitative data lake assessment. *Data & Knowledge Engineering*.

# La fouille de textes en IST : les outils Istex-TDM

**Pascal Cuxac** [1]

*1. INIST - CNRS*
*2 rue Jean Zay, 54500 Vandœuvre lès Nancy*
*pascal.cuxac@inist.fr*

*RÉSUMÉ. L'intelligence artificielle vient bousculer les habitudes des professionnels de tous domaines. À travers l'expérience récente de l'INIST, nous présenterons le développement et la mise en production de nouveaux services dans le domaine de l'Information Scientifique et Technique. Nous faisons le point sur la mise à disposition d'outils de fouille de textes, à destination de non spécialistes, aisément opérables sans connaissances préalables.*

*ABSTRACT. Artificial intelligence is changing the habits of professionals in all fields. Through the recent experience of INIST, we illustrate the development and production of new services in the field of Scientific and Technical Information. We take a look at the availability of text mining tools for non-specialists, which can be easily operated without any prior knowledge.*

*Mots-clés : Fouille de textes ; Intelligence artificielle ; IST ; Publication scientifique ; Offre de service ; Science ouverte*

*KEYWORDS: Text mining ; Artificial intelligence ; STI ; Scientific publication ; Open science*

## 1. Introduction

Les données en libre accès se développent, que ce soit des collections issues de bibliothèques traditionnelles accessibles librement via Internet, mais également des entrepôts numériques regroupant des publications scientifiques nationales ou thématiques : Gallica, Europeana, HathiTrust's digital library, HAL, Isidore, Erudit... De récentes initiatives nationales ont également permis le développement d'archives scientifiques (ISTEX en France, SwissBib en Suisse, GBV en Allemagne, Scholars Portal en Ontario), et nous assistons à la montée en puissance de bases agrégeant des centaines de millions de publication comme CORE ou OpenAlex.

Ces réservoirs de données sont la matière première pour mettre en œuvre des méthodes de fouille de textes qui permettront d'analyser la production scientifique. Cependant, la qualité des données, leur richesse, leur format, sont les premiers écueils rencontrés. Bien entendu des outils existent, mais souvent difficiles à mettre en œuvre par un non spécialiste.

Dans cet article nous illustrons l'utilisation de l'IA dans le domaine de l'IST (Information Scientifique et Technique) à travers les récents développements réalisés à l'INIST[1] et l'offre de service Istex-TDM. Cela sera complété par une démonstration de l'offre, des outils proposés et de leur utilisation.

1. https://www.inist.fr/

## 2. IA et IST : les défis à relever

Les professionnels de l'information doivent répondre aux demandes croissantes de tableaux de bord pour mettre en évidence, entre autres, des taux d'accès ouverts en fonction des disciplines, des instituts ou tout autre indicateur.

Il existe certes des applications « presse bouton » mais leur utilisation dépend d'un abonnement. En plus du coût d'accès, les données ne sont pas toujours homogénéisées, et tous les domaines scientifiques ne sont pas toujours bien représentés, notamment les sciences humaines et sociales (Maddi et De La Laurencie, 2018).

On constate également la mise en ligne croissante, via GitHub ou GitBucket, de programmes permettant de traiter des données. Or leur mise en œuvre souvent complexe n'incite pas les non informaticiens à les utiliser. Des plate-formes comme Cortext, Gargantext et Weka sont aussi disponibles mais elles nécessitent souvent un niveau de connaissance des méthodes de TDM (Text and Data Mining) pour choisir parmi les algorithmes proposés et les paramétrer[2].

Si la fouille de textes a toujours été présente à l'INIST, ce n'est qu'avec le lancement du projet ISTEX[3] que des méthodes d'IA vont être développées pour être appliquées en grandeur nature sur de gros volumes de données, dans un processus industrialisé. Alors que l'IA n'était pas encore un mot-clé passé dans le langage commun, nous avons développé des méthodes d'enrichissement de données à partir notamment de techniques d'apprentissage automatique sous forme de modules intégrés à la chaîne de production (Cuxac et Thouvenin, 2017). Si cette approche a donné de bons résultats, elle a montré un certain nombre de limites : développer et mettre en place un nouveau traitement est un processus complexe à mettre en œuvre, et surtout cela rend très difficile l'utilisation de ces programmes en dehors de la chaîne ISTEX.

Nous nous inscrivons dans le mouvement « Science Ouverte », en publiant tous nos codes, cependant nous voulons aller plus loin en faisant en sorte que qui que ce soit puisse les utiliser, quelque soit ses compétences. Cela doit répondre aux demandes d'utilisateurs, documentalistes ou chercheurs, qui souhaitent pouvoir utiliser ces programmes sur leurs propres données, et en pouvant choisir eux-mêmes les traitements dont ils ont besoin. Le public cible pour ces outils n'est pas le «data scientist», mais plutôt un utilisateur non expérimenté que ce soit en IA, en TDM ou en informatique. C'est un ingénieur ou un chercheur qui souhaite avoir des outils d'aide pour l'analyse de documents ou de corpus, sans avoir à maîtriser des processus complexes.

## 3. L'approche par web-services : d'une IA intégrée dans un processus défini à une boîte à outil modulaire

Nous avons fait le choix de créer et déployer des applications d'IA sous forme de web-services[4] (WS), intégrables dans une chaîne de production comme ISTEX, mais

2. Cortext https://www.cortext.net ; Gargantex https://gargantext.org ; Weka https://waikato.github.io/weka-wiki/
3. https://www.istex.fr/
4. Un WS est une forme spécifique d'API (https://en.wikipedia.org/wiki/Web_service)

également directement utilisables par tout utilisateur désirant traiter ses propres corpus. Ainsi nous passons d'une IA intégrée dans un processus défini à une IA applicable sur ses propres données, avec des contraintes minimales, utilisable par des non spécialistes, et largement extensible pour répondre à de nouveaux besoins.

Les méthodes implémentées, peuvent être complexes, mettant en œuvre des réseaux neuronaux élaborés, avec un nombre élevé de paramètres à optimiser. Afin de faciliter au maximum leur usage, les web-services doivent répondre à un certain nombre d'exigences :

- chaque service ne doit répondre qu'à un seul besoin ;
- il n'y a pas de paramétrage par l'utilisateur ;
- il doit y avoir un seul format d'entrée/sortie très simple ;
- ils doivent être utilisables via l'outil de visualisation Lodex[5].

Les modèles de ML sont construits par des spécialistes TDM, avec l'aide d'experts pour la constitution des corpus d'apprentissage et la validation des algorithmes, puis utilisés par les WS mis à disposition et ainsi applicable aux données bibliographiques, que ce soit sous forme de métadonnées ou de texte intégral (Bonvallot *et al.*, 2022). Pour aider l'utilisateur, le site internet ISTEX-TDM[6] recense les services en production : il permet à la fois d'identifier le service correspondant à ses besoins, connaître son url et avoir une aide sur son utilisation.

A partir de là, le service est utilisable via une interface graphique dans Lodex (outil open source de visualisation de données structurées (Gregorio *et al.*, 2019), (fig 1). Les nouveaux services proposés permettent l'utilisation de méthodes apportant une forte valeur ajoutée aux données traitées sans qu'il soit nécessaire de mobiliser des compétences en informatique, ou datamining.

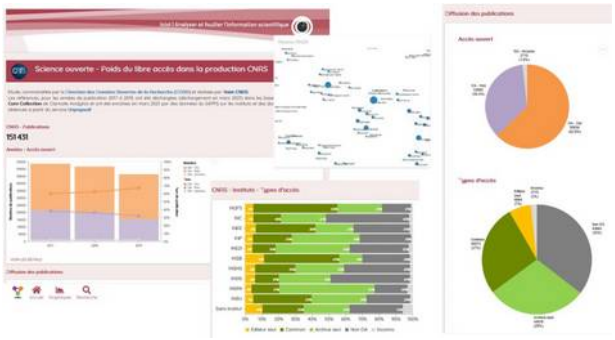



*Figure 1 : Représentations graphiques sous Lodex (d'après Bonvallot et al. 2022).*

Cette nouvelle offre de service est donc là pour répondre à de multiples finalités et s'adresse à tous les professionnels de l'IST qui ont besoin, par exemple, de détecter des thématiques scientifiques, de classer des documents, ou encore de les enrichir pour faire de la bibliométrie. Elle propose des services assez génériques pour être utiles au plus grand nombre, mais est également capable de s'adapter aux

5. https://www.inist.fr/projets/lodex/
6. https://services.istex.fr/

besoins exprimés, et ainsi d'évoluer continuellement pour répondre à de nouveaux usages.

## 4. Conclusions et perspectives

Nous avons mis en place un environnement approprié facilitant le déploiement de services de fouille de textes à partir d'algorithmes d'IA. Cela permet une grande souplesse quant à la modification, l'adaptation ou la création de nouveaux web services. Cette offre de service à destination de non spécialiste de fouille de textes (en priorité appartenant à un établissement de recherche publique), permet de façon extrêmement simplifiée d'exécuter des programmes complexes sans connaissances spécifiques a priori. Par rapport aux plateformes d'analyse de données cette solution est plus légère pour l'utilisateur et facilement interfaçable avec des outils de visualisation.

L'offre de service évolue rapidement, proposant de nouveaux web-services de façon régulière. Très prochainement nous allons mettre à disposition une interface simple permettant à l'utilisateur de charger ses données dans quelques formats simples (y compris csv), de choisir le traitement à faire, et d'être informé par mail avec un lien de téléchargement du résultat quand le traitement est terminé. Notre procédure de mise en production de ces web-services étant automatisé, nous proposons également de travailler avec des chercheurs afin de développer de nouveaux services performants et adaptés aux besoins.

## Bibliographie

Bonvallot V., Parmentier F., Bourguignon L., Clauss I. et Gregorio S. (2022). Le TDM pour tous grâce à des web services au sein de LODEX, outil libre de visualisation, *Revue des Nouvelles Technologies de l'Information, Extraction et Gestion des Connaissances*, RNTI-E-38, 2022, 445-452 (https://editions-rnti.fr/render_pdf.php?p=1002758)

Cuxac P., (2022). L'IA et la fouille de textes à l'INIST : l'IA à portée de tous ?, *Arabesques* 107, 2022, : https://publications-prairial.fr/arabesques/index.php?id=3098)

Cuxac P., Thouvenin N. (2017). Archives numériques et fouille de textes : le projet ISTEX. *Atelier TextMine, conférence EGC*, 24 janvier 2017, Grenoble, France. (https://textmine.sciencesconf.org/data/pages/TextMine17.pdf

Gregorio, S., Collignon A., Parmentier F. et Thouvenin N. (2019). LODEX : des données structurées au web sémantique (https://hal.science/hal-01990444). *Atelier Web des Données ,Conférence EGC, 2019*, Metz, France.

Maddi, A. et De La Laurencie A. (2018). La dynamique des SHS françaises dans le Web of Science : un manque de représentativité ou de visibilité internationale ? (https://hal.science/hal-01922266). working paper.

# *Quo Vadis* INFORSID ?

## Étude des tendances sur la dernière décennie

**Manuele Kirsch Pinheiro**

*Centre de Recherche en Informatique, Université Paris 1 Panthéon Sorbonne*
*90 rue Tolbiac, 97013 Paris, France*
*Manuele.Kirsch-Pinheiro@univ-paris1.fr*

*RÉSUMÉ. La conférence INFORSID constitue la principale conférence dans le domaine des Systèmes d'Information en France (et dans le monde francophone de manière générale). Au cours de son histoire, les travaux qui y ont été présenté témoignent de l'évolution du domaine et de son dynamisme. A l'aube de sa 42ème édition, nous essayons d'identifier dans cet article les tendances apparues lors des 10 dernières éditions. La dernière décennie étant marquée par l'essor des nouvelles technologies et approches ayant un fort impact sur les organisations (e.g. IoT, Big Data, Machine Learning, IA…), nous aimerons connaitre l'influence de ces éléments sur la production scientifique et sur les thématiques abordées dans la conférence. Pour cela, nous adoptons une approche d'analyse volontairement naïve, laquelle se concentre sur les mots les plus fréquemment utilisés dans l'articles, ne se limitant donc pas aux mots clés ou titres des articles.*

*ABSTRACT. The INFORSID conference is the leading conference in the field of Information Systems in France (and in the French-speaking world in general). Over the years, the works presented at the conference bear witness to the evolution of the field and its dynamism. On the eve of its 42nd edition, we attempt in this article to identify the trends that have emerged over the last 10 editions. As the last decade has been marked by the rise of new technologies and approaches with a strong impact on organizations (e.g. IoT, Big Data, Machine Learning, AI, etc.), we wish to know how these elements have influenced scientific production and the themes addressed at the conference. To do this, we are adopting a deliberately naive analysis approach, focusing on the most frequently used terms in the papers, and not limiting ourselves to the keywords or titles of the articles.*

*Mots-clés : INFORSID, sujets d'actualité, tendances, analyse textuelle.*

*Keywords: INFORSID, hot-topics, trends, text analysis.*

### 1. Introduction

Depuis plus de 40 ans, le congrès INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision) réunit chaque année des nombreux chercheurs francophones travaillant sur les Systèmes d'Information. Ces systèmes ont

subi depuis les années des nombreuses transformations, suivant des influences à la fois sociétales et technologiques. Cette dernière décennie a d'ailleurs été marquée par l'essor des nombreuses technologies, comme le *Big Data*, le *Cloud Computing*, l'IoT (*Internet of Things*) ou plus récemment l'Intelligence Artificielle, ayant eu un fort impact sur les organisations et, par extension, sur les Systèmes d'Informations. On peut aujourd'hui parler de Système d'Information Pervasif (Kirsch-Pinheiro et Souveyet, 2019 ; Kirsch-Pinheiro *et al.*, 2023), tellement ces transformations ont permis à ces systèmes de s'étendre au-delà des organisations et notamment sur l'environnement physique.

La question qu'on peut alors se poser est celle de l'influence de ces technologies sur les travaux scientifiques et, plus largement, celle des tendances qu'on pourrait observer dans les recherches menées par la communauté. L'objectif de cet article est ainsi d'essayer d'identifier ces tendances qui auraient influencé la communauté au cours de la dernière décennie. Pour cela, nous proposons une analyse des textes publiés dans les actes des dix dernières éditions du congrès INFORSID. Nous avons choisi une période suffisamment longue pour être significative, mais suffisamment limitée pour permettre une analyse aisée. Cette période a été également marquée par l'essor de nombreuses technologies ayant un fort impact sur les Systèmes d'Information, et donc potentiellement sur les travaux de recherche dans le domaine. Par ailleurs, contrairement aux analyses bibliographiques habituellement réalisées dans ce type d'article (qu'on pourrait appeler « panorama », comme par exemple, Jeyakumaran *et al.* 2021), nous n'avons pas limité notre analyse aux titres, mots clés ou résumé des articles, mais nous avons, au contraire, tenu compte de l'ensemble des textes publiés, réalisant une analyse naïve des termes et des digrammes les plus fréquemment utilisés. L'objectif derrière cette approche d'analyse plutôt naïve est celui d'éviter tout à priori ou parti pris concernant les thématiques. Nous avons ainsi délibérément choisi de se limiter à une simple analyse morpho-syntaxique, s'arrêtant à la classe des mots dans la phrase (verbe, pronom, préposition, conjonction, etc.), sans considérer une quelconque sémantique liée au mot (excluant donc, tout usage d'ontologie). Ce choix vise à éliminer toute interprétation même éventuelle des expressions. Il s'agit donc du texte brut qui est analysé. L'objectif est d'essayer des dénicher derrière le texte brut des articles l'usage des techniques et des technologies qui ne seraient pas centrales à ces travaux (et donc pas forcément mentionnées dans les titres ou dans les mots-clés), mais qui auraient quand-même une influence sur l'ensemble de ces travaux.

Le restant de cet article s'organise donc comme suit : la section 2 détaille la méthode d'analyse utilisée ; la section 3 présente une discussion sur les résultats obtenus à partir de l'analyse, à commencer par une analyse des termes les plus fréquemment utilisés mais également de l'évolution de ceux-ci au cours des années. Puis la section 4 compare certains résultats obtenus avec INFORSID avec ceux d'autres conférences. Enfin, la section 5 présente nos conclusions et perspectives obtenues à partir des analyses.

## 2. Description de la méthode

Dans un article scientifique, toutes les influences et les approches utilisées ne sont pas forcément mentionnées dans le titre, ni même dans les mots-clés. Par exemple, dans Kirsch-Pinheiro *et al.* (2023), on énonce un certain nombre de technologies qui auraient, selon les auteurs, influencé l'apparition des Systèmes d'Information dits Pervasifs. Or aucune de ces technologies n'est mentionnée dans le titre ou les mots-clés, seules quelques-unes sont mentionnées dans le résumé, alors qu'elles ont grandement influencé la notion de SIP défendue par l'article. De même pour Mouysset *et al.* 2019, qui analysent différentes approches de fouille de processus, en pointant en conclusion l'intérêt d'un usage possible de l'Intelligence Artificiel et des Systèmes Multi-Agents. Or aucune des techniques testées, ni les perspectives mises en avant, ne sont mentionnées dans le titre, mots-clés ou même résumé, alors qu'elles sont au cœur même de l'article. On retrouve ainsi des influences, que ce soient technologiques ou techniques, cachées à l'intérieur des articles sans une réelle visibilité par les éléments de référence majeurs de recherche qui sont les titres, les mots-clés et les résumés ou abstracts.

Certes, ces technologies et approches cachées à l'intérieur des travaux ne représentent pas forcément la thématique centrale à ces articles, mais elles en ont une certaine influence. On peut alors légitimement s'interroger sur l'étendue de ces influences, s'il y a des technologies et des techniques qui se détachent, ou même, s'il y a des effets de mode qui auraient pu être observés au cours de cette décennie.

C'est dans cette optique que nous avons mené notre étude. Notre objectif est donc d'analyser l'intégralité des textes publiés dans le congrès INFORSID ces dix dernières années à la recherche des termes le plus fréquemment utilisés afin d'essayer d'identifier à travers ces termes des influences qui pourraient s'y détacher. Pour cela, nous avons, dans un premier moment, téléchargé l'ensemble des actes du congrès INFORSID[1] pour la période 2014-2023. Ensuite, pour chaque document, nous avons procédé de manière suivante (cf. Figure 1) : à l'aide d'un script Python, nous avons extrait le texte des documents PDFs ; Ce texte a été ensuite nettoyé, ce qui comprend l'élimination des caractères spéciaux et de ligature, avant la tokenisation et la lemmatisation du texte avec la bibliothèque SpaCy[2]. Il en résulte un ensemble de mots sans flexions, comme le pluriel ou les temps verbaux. Nous avons alors éliminé de cet ensemble les mots vides (« *stop words* » en anglais[3]) et des expressions couramment utilisées, aussi bien en anglais qu'en français (par exemple, « ainsi », « donc », « aussi »…), ainsi que des abréviations telles que « e.g. », « i.e. » ou encore « etc. ») et certaines classes de mots (verbes, conjonctions, prépositions…) pour ne garder que les termes potentiellement les plus représentatifs du contenu du texte. Enfin, dans chaque document, nous avons calculé, à l'aide de la bibliothèque NLTK[4], les mots et

[1] Disponibles sur http://inforsid.org/actes_conference.php
[2] Disponible sur https://spacy.io
[3] Mots tellement communs qu'ils deviennent non significatifs dans un texte (https://fr.wikipedia.org/wiki/Mot_vide).
[4] https://www.nltk.org

les digrammes les plus fréquents dans le document. Un seuil limite a alors été appliqué, et seuls les mots et digrammes dont la fréquence dépassait ce seuil ont été retenus pour la suite. Nous avons préféré garder un seuil unique pour l'ensemble des années afin qu'une fois réunis, chaque année puisse disposer d'une même proportion de termes dans le corpus finale. Chaque année conserve ainsi son importance dans l'ensemble analysé, malgré les variations dans le nombre d'articles publiés (et donc potentiellement dans le nombre de termes disponibles) qui peuvent exister entre les années. Il s'agit pour nous de conserver une certaine équité entre les différentes éditions de la conférence.

1. Extraction du texte des articles
2. Nettoyage du texte
3. Tokenisation et lemmatisation du texte
4. Suppression des classes des mots non-représentatives
5. Élimination des stopwords, abréviations et expressions idiomatiques
6. Calcul et sélection des mots les plus fréquents
7. Calcul et sélection des digrammes les plus fréquents
8. Export des résultats (CSV, Wordcloud et graphique bar)

*Figure 1. Cycle de traitement pour chaque proceedings.*

L'ensemble des résultats a été alors réunit et analysé, d'abord en fonction de la fréquence globale (toute année confondue), puis en fonction de l'évolution de leur fréquence au cours des années. Deux tours d'analyse ont été réalisés, d'abord avec un seuil de 15 dans un premier passage (seuls les 15 mots et diagrammes les plus fréquents de chaque année ont été retenus), puis avec un seuil de 50 pour un second passage. Ces seuils ont été choisis après plusieurs tests préliminaires avec différentes valeurs : des seuils inférieurs à 15 sont apparus trop stricts alors que ceux supérieurs à 50 ont inclus trop de termes d'usage courante, pas forcément significatifs, rendant l'interprétation des données plus difficile. A chaque tour d'analyse, nous avons observé notamment les mots et digrammes présents dans le premier et le dernier quartile, ainsi que le top 15 des mots et des digrammes les plus fréquents. Les résultats de ces analyses sont discutés dans les sections qui suivent. Les scripts utilisés pour le traitement des documents et l'analyse sont disponibles sur GitHub[5].

## 3. Discussion des résultats

### *3.1. Analyse par fréquence*

Notre première analyse concerne les termes utilisés de manière fréquente lors des dix dernières éditions du congrès INFORSID. Les deux tours d'analyse que nous avons réalisés (cg. Figure 2) nous ont permis d'identifier des termes les plus fréquemment utilisées lors de ces éditions. Lors du premier tour, les 15 mots les plus fréquents de chaque année ont été retenus, ce qui nous offert un ensemble de 62 mots

[5] https://github.com/mkirschpin/proceedingsanalysis

au total, avec une fréquence moyenne par année de 212 apparitions (max de 747 et min de 92, avec un écart-type de 117,5). Lors du second round, un total de 183 mots différents a été retrouvé, avec une fréquence moyenne par année de 133 apparitions (min 53 et écart-type de 85). Lorsque nous réunissons l'ensemble des données obtenues à chaque tour, nous obtenons une fréquence moyenne de 773 (seuil 15) et de 541 (seuil 50), avec une fréquence maximale à 4197 et minimale à 53 (seuil 50).

| | count | year |
|---|---|---|
| count | 150.000000 | 150.000000 |
| mean | 212.726667 | 2018.500000 |
| std | 117.469110 | 2.881904 |
| min | 92.000000 | 2014.000000 |
| 25% | 136.500000 | 2016.000000 |
| 50% | 179.000000 | 2018.500000 |
| 75% | 241.250000 | 2021.000000 |
| max | 747.000000 | 2023.000000 |

(a)

| | count | year |
|---|---|---|
| count | 500.000000 | 500.000000 |
| mean | 133.392000 | 2018.500000 |
| std | 85.784266 | 2.875158 |
| min | 53.000000 | 2014.000000 |
| 25% | 83.000000 | 2016.000000 |
| 50% | 112.000000 | 2018.500000 |
| 75% | 147.000000 | 2021.000000 |
| max | 747.000000 | 2023.000000 |

(b)

*Figure 2. Fréquences obtenues lors de deux rounds d'analyse avec (a) un seuil de fréquence de 15, et (b) un seuil de fréquence de 50.*

Dans le dernier quartile, comportant les termes les plus fréquemment utilisés toute année confondue, nous retrouvons logiquement des indicateurs des thématiques majeures de notre communauté (voir Figure 3), dont notamment les termes « *donnée* » (mentionné plus de 4000 fois dans l'ensemble de documents), « *information* » et « *modèle* » (respectivement à plus 3000 et 2000 mentions). On observe également la présence, parmi les plus fréquent, des mots « *processus* » et « *utilisateurs* ». Le premier peut être vu comme un marqueur des thématiques liées à la gestion de processus en entreprise. Cette analyse est corroborée par l'analyse des digrammes les plus fréquents (cf. Figure 4), où on retrouve différentes expressions liées en particulier à la notion de processus et à la fouille de processus (e.g. « *processus*, *métier* », « *modèle, processus* », « *process, model* », « *process, mining* »). On observe ainsi l'importance des travaux autour des processus métiers et de leur modélisation pour la communauté (p.ex. Ismaïli-Alaoui *et al.* 2022 ; Biard *et al.* 2017), mais également l'essor des travaux sur la fouille des processus. Apparus sur les actes d'INFORSID 2019, les digrammes autour de ce sujet cumulent une fréquence de 165 mentions, se trouvant tous sur le quartile supérieur, avec une mention spéciale pour le diagramme « *van*, *der* ». Celui-ci cumule à lui seul 59 mentions (largement supérieur à la moyenne de 35 obtenue dans l'échantillon de top 15), dont 53 font référence à l'auteur M.W. van der Aalst, considéré comme un des précurseurs du *process mining* (van der Aalst *et al.* 2004 ; van der Aalst, 2012).

Le mot « *utilisateur* », quant à lui, peut être perçu comme un indicateur de l'importance accordée par la communauté à l'utilisateur et à la place centrale qu'il occupe dans les Systèmes d'Information. Ceci se confirme par la présence

d'expressions comme « *profil*, *utilisateur* » dans les diagrammes les plus fréquents (Figure 4), ainsi que par la présence d'autres mots connexes parmi celles mentionnées au premier quartile (représentant les mots retenus les moins fréquents, illustrés par la Figure 5) : « *acteur* », « *communauté* », « *personne* », « *psychologique* ». Nous pouvons citer, par exemple, les travaux de Bour *et al.* (2019), Arduin (2021), ou encore Zhong et Negre (2021).

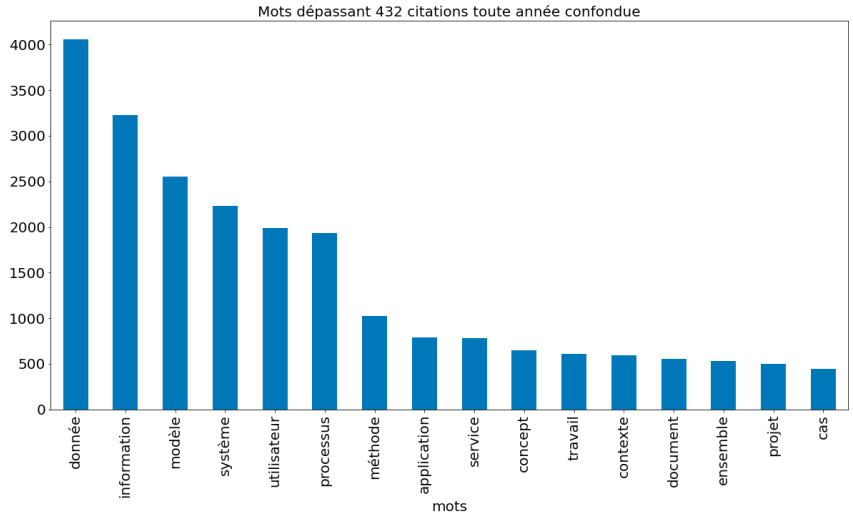



*Figure 3. Mots les plus fréquents (dernier quartile) toute année confondue (seuil de fréquence de 15 par année).*

On peut toutefois s'étonner de l'absence de certains mots associés à cette vision « centrée utilisateur » et caractéristiques de certaines thématiques bien connues du domaine comme les mots « *ingénierie* » et « *exigence* ». Le mot « *ingénierie* » est absent du top 50 des mots les plus fréquents, alors que le mot « *exigence* » apparaît uniquement sur trois années de l'échantillon (2015, 2016 et 2022). Du même, l'expression « *ingénierie*, *exigence* » apparait dans les plus fréquents de l'année 2015 uniquement. Cependant, cette absence ne doit pas être interprétée comme l'absence de cette thématique aux dernières éditions d'INFORSID, comme peuvent le témoigner les travaux de Ponsard *et al.* (2015), Kang et Saint-Dizier (2015) et Chan *et al.* (2013), mais simplement comme une présence plus discrète de ces mots comparativement à ceux utilisés par d'autres thématiques.

Une analyse attentive de la Figure 2 met en avant la présence, sans surprise, des mots typiquement associés aux thématiques les plus traditionnelles abordées par la communauté, comme la gestion de processus déjà mentionnée ; la modélisation, à travers des mots tels que « *modèle* », « *méthode* » ou « *concept* » et les expressions « *langage*, *modélisation* », « *modèle*, *méthode* » ou encore « *modèle*, *données* » ; et la gestion de projet à travers des mots comme « *projet* » ou « *scrum* » (mentionné parmi

les moins fréquents dans le top 50) et les digrammes « *projet*, *agile* », « *méthode*, *agile* », et « *pratique*, *agile* ». Les digrammes les plus fréquents (en tenant compte de toutes les éditions considérées) représentent d'ailleurs assez bien la communauté puisqu'il s'agit des expressions « *système*, *information* » et « *base*, *donnée* ».

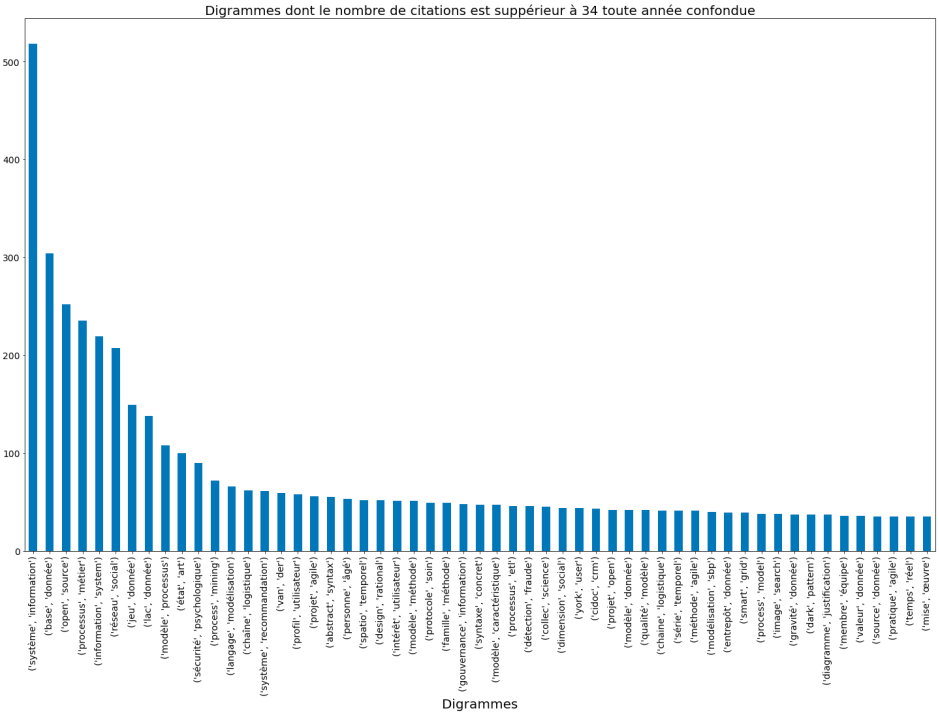



*Figure 4. Digrammes les plus fréquents (dernier quartile) toute édition confondue (seuil annuel de fréquence fixé à 15).*

Le troisième digramme parmi les plus fréquents est un peu plus étonnant. Il s'agit de l'expression « *open*, *source* », mentionnée plus de 267 sur les éditions 2016, 2018 et 2019. Cette présence pourrait être interprétée comme un indicateur de la volonté de la communauté d'aller vers une science livre et accessible. On peut citer des travaux comme celui de Viseur (2016), sur la gouvernance de projets open source. Mais elle pourrait également être le simple indicatif d'un usage récurrent de logiciels suivant ce modèle, comme c'est le cas de Gillet *et al.* (2019).

Enfin, concernant les mots les moins fréquents retenus dans notre étude (cf. Figure 5 et Figure 6), nous pouvons observer l'apparition de certains mots représentatifs de technologies et de techniques, comme « *blockchain* », « *scrum* » ou encore « *bert* », référent au modèle de langage en NLP (Gillioz *et al.* 2020). Des références à ces technologies sont également visibles parmi les 50 digrammes les moins fréquents (cf. Figure 7), où on peut apercevoir les expressions « *modèle, bert* », « *architecture,*

*transformer* », et « *machine, learning* ». On peut également souligner la présence dans ces figures des mots « *rgpd* » (Figure 5), « *attaque* » et « *confiance* » (Figure 6), ou encore le digramme « *exigence*, *sécurité* » (Figure 7), dénotant l'intérêt grandissant par la communauté par les aspects de conformité et sécurité.

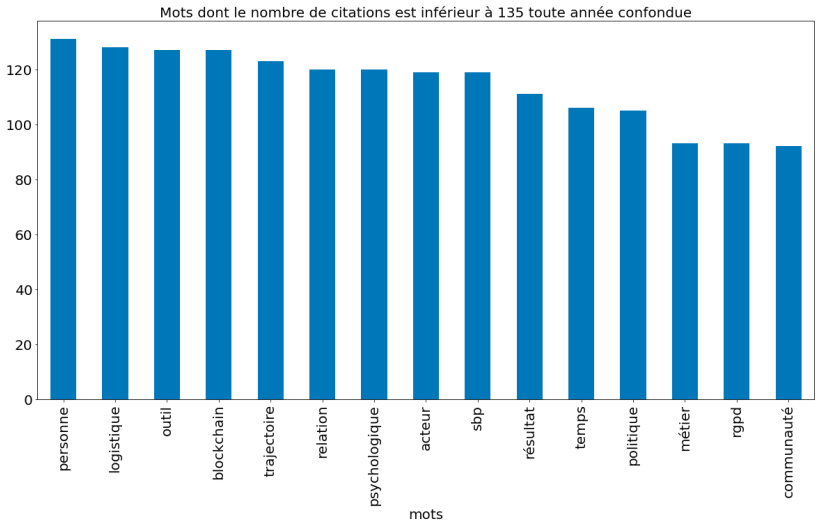



*Figure 5. Mots dont la fréquence totale (toute année confondue) est inférieure au seuil du premier quartile (seuil de fréquence annuelle défini à 15).*

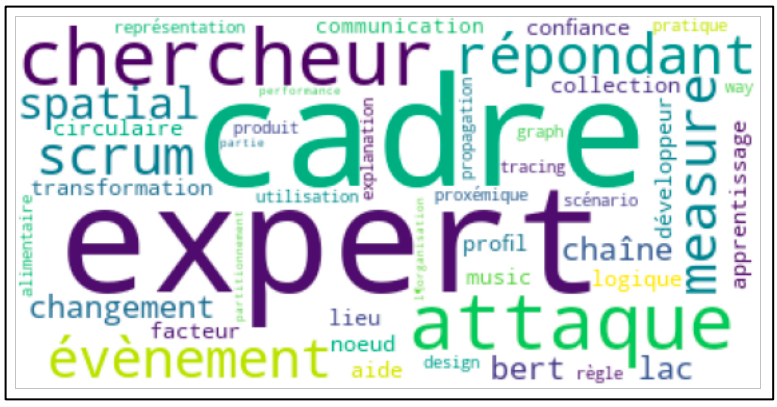



*Figure 6. Nuage de mots le moins fréquents retenus utilisant un seuil de fréquence annuel à 50 unités.*

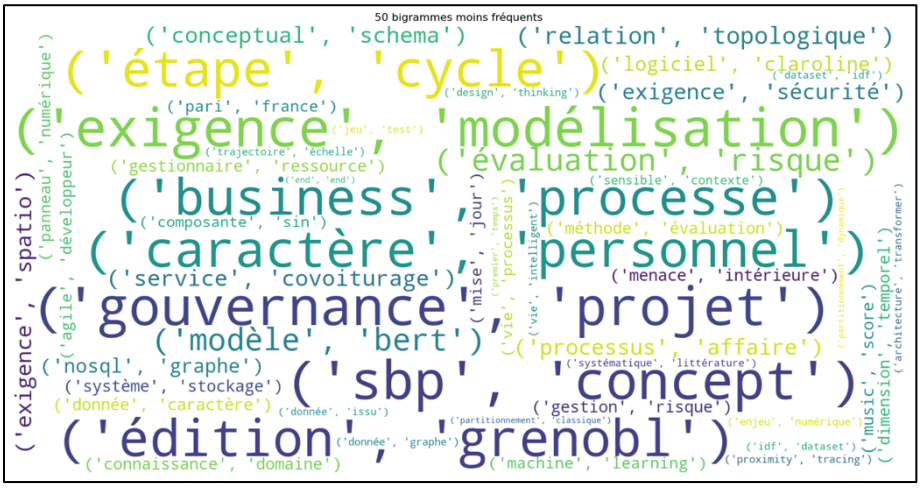



*Figure 7. Cinquante digrammes les moins fréquents (seuil de fréquence annuelle défini à 50).*

### *3.2. Évolution au cours des années*

Après avoir considéré les mots les plus fréquents dans l'ensemble des éditions étudiées, nous avons considéré l'évolution de cette fréquence au cours des différentes éditions du congrès INFORSID. En effet, la fréquence de certains mots varie en fonction des années, disparaissant du top 15 (ou même du top 50) à certaines éditions. La Figure 8 illustre ces variations pour les 10 mots les plus fréquemment mentionnés au cours de la décennie. On peut observer la présence régulière de certains termes parmi les plus fréquents, comme « *donné* », « *information* » ou encore « *utilisateur* ». La présence constante de ce dernier conforte la vision d'une communauté jusqu'ici particulièrement concernée par les utilisateurs. Le même peut être constaté au niveau des digrammes (cf. Figure 9), dont la régularité de certaines expressions clés, comme « *base*, *donnée* », « *système*, *information* » et « *processus, métier* » contraste avec les variations observables sur des expressions comme « *système, recommandation* », « *process, mining* » et « *lac, donné* ». Ces variations peuvent être symptomatiques de la popularité de certains thématiques à certains moments.

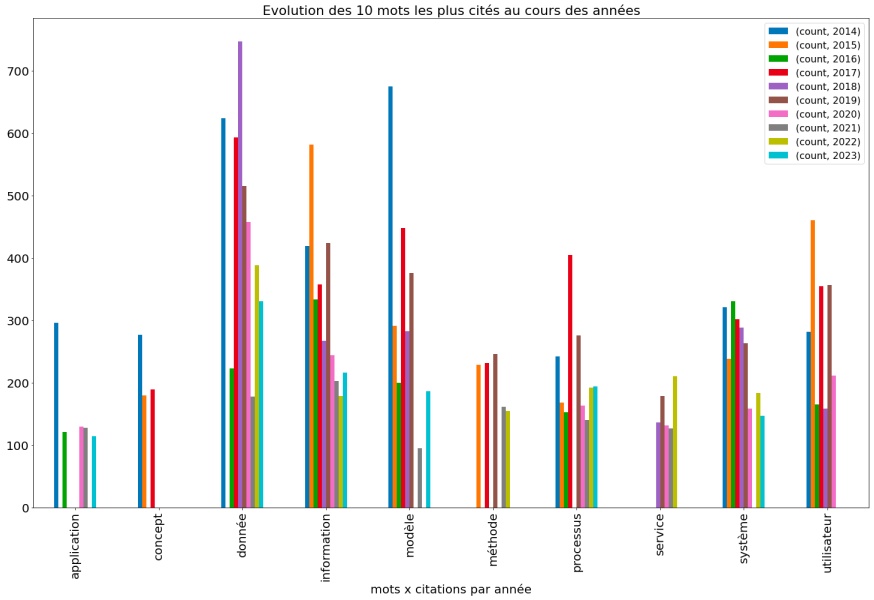



*Figure 8. Variations dans la fréquence des mots au cours de 10 dernières années (seuil de fréquence annuel utilisé : 15).*

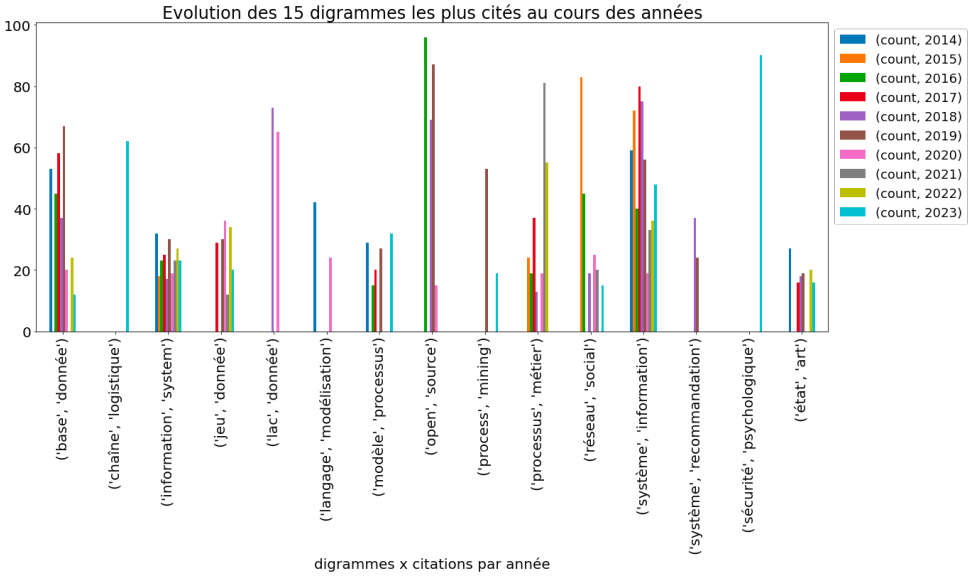
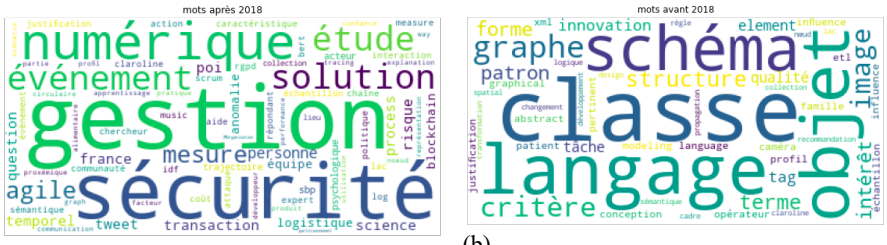


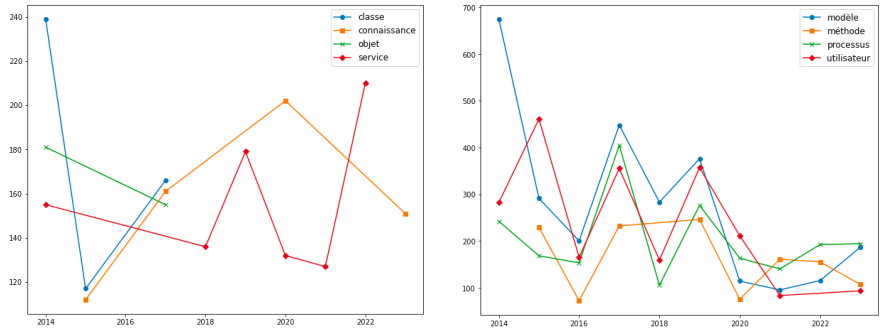

*Figure 9. Digrammes les plus fréquents au cours de la décennie (seuil de fréquence annuel utilisé : 15).*




(a) (b)

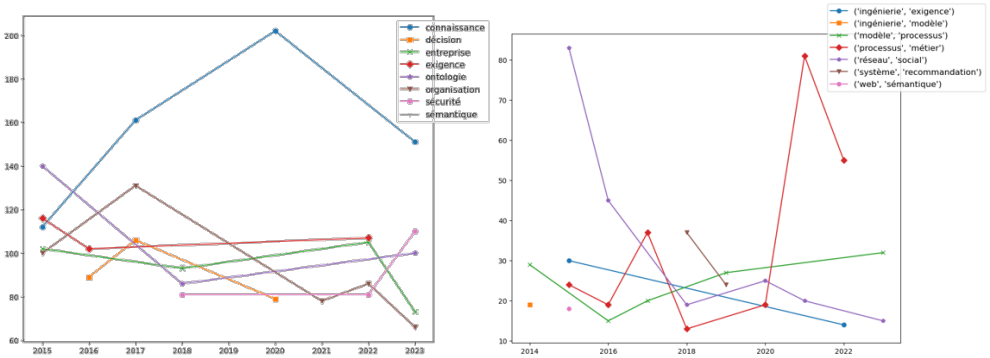
*Figure 10. Nuage de mots contenant en (a) les mots présents dans les top 50 annuels uniquement après 2018 ; et en (b) les mots présents dans les top 50 annuels uniquement avant 2018.*

Afin de mieux comprendre ces variations, nous avons regardé plus en détails les mots apparaissant dans la liste des plus fréquents uniquement avant et après 2018 (Figure 10). Nous pouvons observer dans le nuage de mots représenté dans la Figure 10a la présence de technologies dont l'essor est plus récent (comme « *blockchain* » et « *bert* »), mais également des mots comme « *sécurité* », « *rgpd* » ou encore « *tweet* », correspondant à des préoccupations majeures qui ont gagné en intensité ces dernières années. Inversement, le nuage de mots représenté Figure 10b indique les mots présents dans le top 50 des mots plus fréquents uniquement avant 2018. On observe notamment la présence des mots « *objet* » et « *classe* », ce qui laisse penser à une certaine perte de vitesse sur les sujets de recherche concernant ces mots. La Figure 11a conforte cette idée à travers une représentation de la fréquence des mots « *classe* » et « *objet* » au cours des années. Ceux-ci sont comparés au mot « *service* », présent dans Figure 8, et au mot « *connaissance* », présent en 56$^{ème}$ position des 500 mots les plus fréquemment utilisés (avec le seuil de fréquence annuel de 50). Ces sont des termes assez significatifs de certaines thématiques de recherche en Système d'Information, comme l'orientation service et la représentation de connaissances, tout comme les termes « *classe* » et « *objet* », typiques de sujets comme l'ingénierie de modèles. On observe alors que ceux derniers, « *classe* » et « *objet* », ne sont plus dans la liste de top 50 de mots les plus fréquents depuis 2017, alors que les deux autres y sont régulièrement mentionnés. Du même, lorsqu'on regarde dans le détail certains mots mentionnés dans la Figure 8 et la Figure 3 comme étant fréquemment utilisés, on observe une certaine perte de vitesse pour des mots comme « *modèle* », ce qui vient quelque part confirmer l'observation réalisée avec les mots « *classe* » et « *objet* ».




(a) (b)

*Figure 11. Variation sur les fréquences de certains mots dans le top 50 au cours de ces 10 dernières années.*




(a) (b)

*Figure 12. Variation sur les fréquences de mots et de digrammes utilisés dans les noms des sessions (seuil de fréquence utilisé : 50).*

Enfin, certains de ces mots (Figure 11) sont régulièrement utilisés en tant que thématique privilégiée ou dans les sessions de la conférence. Nous avons alors voulu observer l'évolution de certains de ces mots (Figure 12a) et digrammes (Figure 12b) régulièrement utilisés dans les sessions de ces 10 dernières éditions (extraits à partir des actes de ces éditions). Parmi les digrammes (Figure 12b), on peut observer notamment le diagramme « *réseau, social* », apparu en 2015 mais dont la fréquence s'est réduite depuis 2018, mais également les digrammes « *processus, métier* » et « *modèle, processus* » qui se développent depuis 2018 (peut-être grâce à l'essor des recherches autour de la fouille de processus). Par rapport aux mots couramment utilisés dans les titres des sessions, on voit que leur fréquence est globalement stable, avec une curieuse réduction sur l'usage des termes « *organisation* » et « *entreprise* » ces dernières années. On peut alors s'interroger si la vision orientée organisation typique des SI ne serait pas en train de céder la place à une vision plus sociétale.

## 4. Comparatif avec d'autres événements

Après avoir étudié les termes et digrammes les plus fréquents sur les 10 dernières années d'INFORSID, nous avons voulu comparer les résultats obtenus avec d'autres conférences semblables. Pour cela, nous avons comparé nos résultats avec les termes et digrammes obtenus à partir des conférences RCIS[6] et PERCOM[7]. La première est une conférence internationale en Systèmes d'Information, avec des nombreux auteurs en commun avec la communauté INFORSID, alors que la seconde est une conférence internationale en Informatique Pervasive, de taille comparable à INFORSID (20 à 40 articles publiés par année).

| word | count |
|---|---|
| donnée | 4197 |
| information | 3226 |
| modèle | 2783 |
| système | 2316 |
| utilisateur | 2166 |
| processus | 2038 |
| méthode | 1277 |
| application | 1266 |
| cas | 1218 |
| travail | 1119 |

(a)

| word | count |
|---|---|
| model | 11284 |
| process | 10711 |
| information | 8506 |
| user | 6325 |
| system | 5341 |
| used | 5048 |
| research | 4985 |
| hav | 4702 |
| using | 4567 |
| work | 4240 |

(b)

| word | count |
|---|---|
| device | 4879 |
| using | 4863 |
| tim | 4301 |
| model | 4081 |
| user | 3777 |
| used | 3272 |
| application | 2918 |
| information | 2738 |
| set | 2652 |
| signal | 2621 |

(c)

*Figure 13. Comparaison entre les 10 mots les plus fréquents des conférences INFORSID (a), RCIS (b) et PERCOM (c).*

Lorsqu'on observe les mots les plus fréquents (cf. Figure 12), nous voyons que RCIS et INFORSID partagent plusieurs mots (« *modèle* », « *processus* », « *utilisateur* »…), alors que PERCOM voit apparaitre des mots davantage liées aux technologies, comme « *device* » et « *signal* ».

Cependant, l'observation des digrammes (voir Figure 14) permet de mieux observer les différences entre les communautés. Alors que les digrammes les plus fréquents de RCIS (cf. Figure 14a) font essentiellement référence, comme pour INFORSID (cf. section 3), aux fondamentaux de la communauté (Figure 14a) comme « *business, process* », « *process, mining* » et « *process, model* », les diagrammes plus fréquents à PERCOM démontrent la connotation technologique de la communauté d'Informatique Pervasive, avec des termes comme « *machine, learning* », « *deep, learning* », *« neural, network* », « *transfer, learning* », « *tim, serie* » (pour « *time series* »), qui montrent clairement l'influence de l'IA dans la communauté. Même si l'IA n'est pas une thématique centrale à la communauté PERCOM, on peut quand-même observer l'importance de ces techniques pour la communauté, ce qui n'est pas le cas pour la communauté de Systèmes d'Information, même si on peut observer la présence de l'expression « *machine, learning* » sur les digrammes les plus fréquents

[6] Research Challenges in Information Science (https://www.rcis-conf.com)
[7] Int. Conference on Pervasive Computing and Communications (https://www.percom.org)

de RCIS. Ceci peut s'expliquer, entre autres, par la présence croissante de travaux sur la fouille de processus sur cette conférence, comme l'attestent la présence sur la Figure 14 des digrammes « *process, mining* », « *event, log* » et « *process, model* », tous en rapport avec cette thématique.


*Figure 14. Les 50 digrammes les plus fréquents à RCIS (a) et à PERCOM (b).*

## 5. Conclusions et perspectives

Cet article a présenté une analyse naïve, par fréquence des mots, de l'ensemble des actes du congrès INFORSID au cours de ces dix dernières années. Même s'il reste difficile de tirer des conclusions sur les tendances abordées par la communauté, nous avons pu observer une communauté très ancrée sur ses fondamentaux, dont le cœur de métier demeure les Systèmes d'Information. On peut également constater une communauté concentrée sur des thématiques de recherche majeures, qui n'a pas cédé aux effets de modes concernant les technologies, puisque les mots ressortis dans nos analyses sont majoritairement représentatifs des thématiques et non des technologiques ou techniques dont il a été beaucoup question ces dernières années.

Même l'essor significatif de l'Intelligence Artificielle sur ces différentes formes n'est pas vraiment visible à travers les mots les plus fréquents utilisés ces dernières années.

On peut cependant regretter l'absence de certains mots liés notamment aux préoccupations environnementales et de sustainabilité. Même si l'intérêt pour ces problématiques est monté ces dernières années, elles semblent encore marginales face à d'autres sujets plus traditionnels dans le domaine des Systèmes d'Information

Enfin, les observations réalisées dans cet article invitent à davantage d'exploration. Parmi les perspectives envisagées, nous pouvons souligner l'usage d'ontologies, afin d'avoir une première couche d'interprétation sur les termes observés et l'analyse de trigrammes (à la recherche d'expressions plus complexes), mais également l'usage d'autres algorithmes comme TD-IDF[8] et t-SNE[9] ou encore de l'ACF, pour l'analyse de la communauté d'auteurs (Jaffal, 2019).

*Remerciements*

*L'auteur remercie chaleureusement ses collègues Pr. Bénédicte Le Grand, de l'Université Paris 1 Panthéon Sorbonne, et Pr. Luiz Angelo Steffenel, de l'Université Reims Champagne-Ardenne, pour tous les échanges particulièrement constructifs qui ont permis cet article de voir le jour.*

## Bibliographie

Biard, T., Bourey, J.-P., Bigand, M. (2017). DMN (Decision Model and Notation) : De la Modélisation à l'Automatisation des Décisions. *Actes du XXXVème Congrès INFORSID*, Toulouse, France, May 30 - June 2, 2017

Bour, R., Vallès-Parlangeau, N., Soule-Dupuy, C. (2019). DEMOS : une méthode de conception participative pour un empowerment démocratique des utilisateurs de SI. *Actes du XXXVIIème Congrès INFORSID*, Paris, France, June 11-14, 2019

Chan, A., Fernandes Pires, A., Polacsek, T. (2023). Éliciter, raffiner et attribuer des buts aux bons acteurs à partir d'objectifs de haut niveau, *Actes du XLI Congrès INFORSID*, La Rochelle, France, May 30 - June 2, 2023.

Gillet, A., Leclercq, E., Cullot, N. (2019). Lambda Architecture pour une analyse à haute performance des données des réseaux sociaux. *Actes du XXXVIIème Congrès INFORSID*, Paris, France, June 11-14, 2019

Gillioz, A., Casas, J., Mugellini, E., Khaled, O. A. (2020). "Overview of the Transformer-based Models for NLP Tasks," 2020 15th Conference on Computer Science and Information Systems (FedCSIS), Sofia, Bulgaria, 2020, pp. 179-183, doi: 10.15439/2020F20

Ismaïli-Alaoui, A., Benali, K., Baïna, K. (2022). Traitement des événements complexes pour une gestion proactive des instances d'un processus métier, *Actes du XLème Congrès INFORSID*, Dijon, France, May 31 - June 3, 2022.

[8] *Term Frequency-Inverse Document Frequency* : https://fr.wikipedia.org/wiki/TF-IDF
[9] *t-Distributed Stochastic Neighbor Embedding* : https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding

Jaffal, Ali. (2019). Aide à l'utilisation et à l'exploitation de l'Analyse de Concepts Formels pour des non-spécialistes de l'analyse des données. Thèse de doctorat en Informatique. Université Paris 1 Panthéon Sorbonne. https://hal.science/tel-02526323v1

Jeyakumaran, S., Rychkova, I., Deneckere, R. (2021). Comment la Conformité au RGPD est intégrée dans les Pratiques de Gestion de Processus Métier (BPM) ?: Une revue systématique de la littérature, *Actes du XXXIXème Congrès INFORSID*, Dijon, France, June 1-4, 2021.

Kang, J., Saint-Dizier, P. (2015). Une expérience d'un déploiement industriel de LELIE: une relecture intelligente des exigences, *Actes du XXXIIIème Congrès INFORSID*, Biarritz, France, May 26-29, 2015.

Kirsch Pinheiro, M., Roose, P., Steffenel, L. A., Souveyet, C. (2023). What Is a "Pervasive Information System" (PIS)? In: Kirsch Pinheiro, M., Souveyet, C., Roose, P., Steffenel, L. A., (Eds.), *The Evolution of Pervasive Information Systems*. Springer International Publishing: Cham, 2023; pp 1–17. https://doi.org/10.1007/978-3-031-18176-4_1.

Kirsch-Pinheiro, M., Souveyet, C. (2019). Le Rôle des ressources dans l'évolution des Systèmes d'information. In *Actes du XXXVIIème Congrès INFORSID, Paris, France, June 11-14, 2019*; pp 85–97.

Mouysset, F., Picard, C., Bortolaso, C., Migeon, F., Gleizes, M.-P., Maurel, C., Derras, M. (2019). Investigations of Process Mining Methods to discover Process Models on a Large Public Administration Software, *Actes du XXXVIIème Congrès INFORSID*, Paris, France, June 11-14, 2019.

Ponsard, C., Darimont, R., Michot, A. (2015). Combining Models, Diagrams and Tables for Efficient Requirements Engineering: Lessons Learned from the Industry, *Actes du XXXIIIème Congrès INFORSID*, Biarritz, France, May 26-29, 2015.

van der Aalst W. M. P., Weijters T., Maruster L. (2004). Workflow Mining: Discovering Process Models from Event Log. Transactions on Knowledge and Data Engineering, vol. 16, no. 9, pp. 1128-1142, IEEE.

van der Aalst W. (2012). Process mining: Overview and opportunities. ACM Trans. on Management Information Systems, vol. 3, no 2, p. 7.

Viseur, R. (2016). Gouvernance des projets open source, *Actes du XXXIVème Congrès INFORSID*, Grenoble, France, May 31 - June 3, 2016

Zhong, J., Negre, E. (2021). AI: To interpret or to explain? *Actes du XXXIXème Congrès INFORSID*, Dijon, France, June 1-4, 2021

# Amélioration d'une Méthode de Clustering des Traces Moodle via l'Encodage des SSF

**Noura Joudieh** [1], **Marwa Trablesi** [1], **Ronan Champagnat** [1], **Mourad Rabah** [1], **Samuel Nowakowski** [2], **Nikleia Eteokleous** [3]

1. *L3i - Université de La Rochelle*
   *Avenue Michel Crépeau*
   *17 042 La Rochelle, France*
   *nom.prenom@univ-lr.fr*
2. *LORIA - Université de Lorraine*
   *Campus Scientifique, 615 rue du jardin-botanique*
   *54 506 Vandœuvre-lès-Nancy, France*
   *samuel.nowakowski@loria.fr*
3. *Frederick University*
   *Department of Education*
   *3080 Limassol, Cyprus*
   *n.eteokleous@frederick.ac.cy*

RÉSUMÉ. *Les apprenants mettent en place des stratégies d'apprentissages variées, ce qui rend leur traces d'apprentissage riches et précieuses pour déterminer des recommandations de parcours d'apprentissage pour d'autres apprenants. Dans ce contexte, la fouille de processus permet de découvrir des modèles qui révèlent les parcours d'apprentissage des apprenants dans une plateforme éducative. Dans cet article, nous discutons des limitations et proposons des améliorations d'une approche de regroupement des traces nommée « FSS-encoding ». Nos améliorations visent à enrichir la définition du vecteur caractérisant les traces. Notre méthode a été appliquée aux traces Moodle collectées entre 2018 et 2022 à l'Université Frederick à Chypre.*

ABSTRACT. *Learners adopt various learning patterns and behaviors while learning, rendering their experience a valuable asset for recommending learning paths for other learners. Process Mining is useful in this case to discover models that reveal learners' taken learning paths in an educational platform. In this paper, we address the limits of and improve on a feature-based trace clustering approach known as FSS-encoding. Our enhancements include a refined pattern selection, preserving the uniqueness of less frequent events and increasing the overall effectiveness of the trace clustering process. Our method was applied to Moodle logs collected from 2018 to 2022 in the Frederick University.*

MOTS-CLÉS : *SI pédagogique, Scénarios d'apprentissage, trace clustering*

KEYWORDS: *Learning management system, Trace Clustering in process mining, Learning Paths*

## 1. Introduction

L'évolution des technologies a étendu les possibilités d'apprentissage au-delà des salles de classe traditionnelles et des interactions conventionnelles entre enseignants et élèves. De nos jours, l'apprentissage en ligne est de plus en plus populaire, offrant un accès à une multitude de ressources éducatives à tout moment et en tout lieu. Cependant, cette accessibilité accrue comporte son lot de défis : les apprenants peuvent se sentir surchargés lorsqu'ils cherchent à atteindre leurs objectifs d'apprentissage, ce qui peut potentiellement diminuer leur motivation et leur efficacité.

Dans cette perspective, les systèmes de recommandation (RS) (Aggarwal, 2016) pour l'apprentissage en ligne visent à personnaliser l'expérience d'apprentissage en filtrant de manière intelligente le contenu en ligne en fonction des préférences, des actions et des besoins individuels des apprenants, s'éloignant ainsi des modèles génériques. De plus, les utilisateurs des systèmes d'information laissent des traces enregistrées par le système de journalisation sous forme de journaux d'événements.

La fouille de processus (*Process Mining*), une discipline qui combine la fouille de données, l'apprentissage automatique et la modélisation des processus métier, exploite ces journaux d'événements pour découvrir les modèles de processus qui décrivent le comportement des utilisateurs au sein d'un système (Aalst, 2016). Dans les plates-formes et systèmes éducatifs, cela a ouvert la voie à de prometteurs travaux de recherches visant à identifier les comportements des étudiants lorsqu'ils s'engagent dans diverses activités d'apprentissage telles que suivre un cours ou passer une évaluation (Cenka, Anggun, 2022).

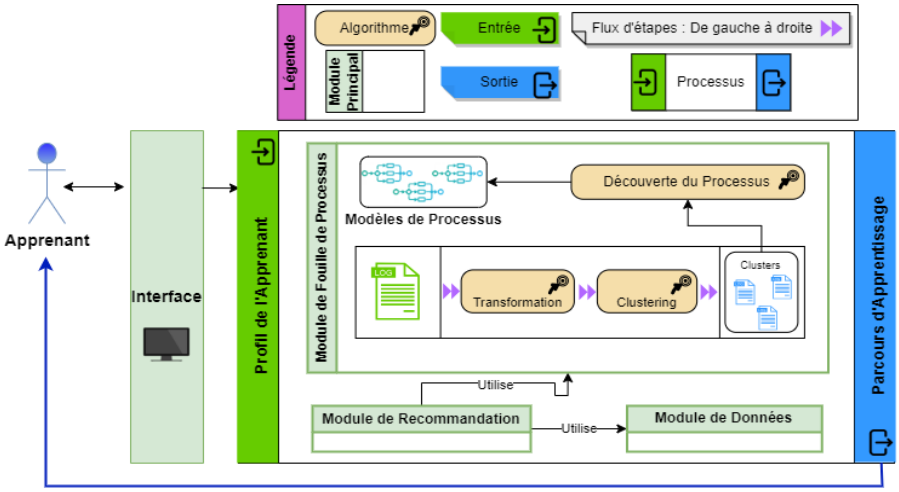



FIGURE 1 – Plate-forme pour la recommandation de parcours d'apprentissage

Dans nos travaux précédents (Joudieh *et al*., 2023) et comme présenté dans la Figure 1, nous avons proposé un framework pour recommander un parcours d'apprentissage adaptatif personnalisé pour un apprenant possédant un objectif d'apprentissage, en utilisant son parcours d'apprentissage passé extrait via la fouille de processus. Ce framework est composé de trois modules principaux : le module de recommandation, le module de données et le module de *Process Mining*. Le présent article se concentre sur le module de fouille de processus. Ce dernier est chargé d'analyser les logs du système de gestion de l'apprentissage Moodle pour découvrir un modèle montrant les parcours d'apprentissage empruntés par les étudiants. Ces modèles forment ainsi la base de recommandations de parcours d'apprentissage personnalisées et efficaces. Cependant, une grande quantité de traces peut générer des modèles incompréhensibles en forme de « spaghettis », d'où la nécessité du regroupement des traces (*trace clustering*) (Song *et al*., 2008) comme étape préliminaire pour identifier les différents types de parcours d'apprentissage et découvrir ainsi des modèles plus explicites.

Guidés par cet objectif, nous adoptons une approche récente de regroupement des traces (Trabelsi *et al*., 2021) qui repose sur l'encodage des traces en termes de sous-séquences fréquentes, nommée *FSS Encoding* (*Frequent Subsequent Patterns Encoding*). Cette méthode générique, intialement développée pour les utilisateurs des bibliothèques numériques, a réussi à identifier trois profils distincts d'utilisateurs à partir de données simulées, confirmés par des données réelles de la Bibliothèque Nationale de France. La méthode a montré que l'utilisation répétée d'une même séquence d'événements pour une tâche indique que cette séquence satisfait l'utilisateur pour atteindre son objectif. En éducation, il est intéressant de prendre en compte les patterns d'apprentissage (une sous-séquence) plutôt que les activités atomiques en analysant les traces d'apprentissage. Une simple consultation de ressources peut être informative, mais un pattern construit d'une consultation de ressources suivie d'exercices est bien plus intéressant. Pour cela, une méthode de regroupement basée sur les FSS, peut révéler les stratégies d'apprentissage et permettre de regrouper les étudiants en conséquence.

Dans le présent article, nous abordons les limites de cette méthode et proposons une extension qui améliore les résultats de regroupement et les modèles découverts. L'approche améliorée est appliquée aux traces Moodle collectées auprès de 471 étudiants suivant des cours au Département d'informatique et de génie informatique de l'Université Frederick à Chypre pour la période 2018-2022.

Dans la section 2, nous décrivons les différentes approches de regroupement des traces, suivi d'une analyse approfondie de la méthode *FSS Encoding*, de ses limitations et des améliorations proposées dans la section 3. La section 4 fournit une description détaillée de la collecte des logs générés par la plateforme Moodle et du prétraitement des données. Les résultats de l'application de la méthode FSS améliorée sur les logs Moodle sont présentés dans la section 5, ainsi que la compairison avec la méthode originale et une autre méthode identifiée dans l'état de l'art. La section 6 conclut le travail et présente les perspectives.

## 2. État de l'art

Le *Process Mining* est un domaine scientifique qui comble le fossé entre l'analyse orientée données et l'analyse orientée processus, visant à extraire des connaissances à partir des journaux d'événements (logs). Les techniques de fouille de processus sont appliquées dans divers domaines par les hôpitaux, les banques ou encore les collectivités territoriales (Aalst, 2016 ; Lu *et al*., 2019 ; Trabelsi *et al*., 2021).

Ces techniques utilisent les logs comme entrée pour générer, améliorer ou valider des modèles de processus (Aalst, 2016). Un journal d'événements se compose d'un ensemble de traces d'exécution, chacune représentant une instance spécifique du processus. Prenons par exemple le flux de travail d'un étudiant sur une plateforme d'apprentissage en ligne. À partir de l'action de consulter un cours (*Course viewed*), l'étudiant peut naviguer pour explorer des éléments spécifiques dans le cours (*Course module viewed*). Ces éléments peuvent inclure des cours, des vidéos, des devoirs ou des quiz. En sélectionnant un élément de quiz, l'étudiant procède ensuite à la soumission de ses réponses (*Quiz submitted*). Chacune de ces activités constitue une trace unique au sein du processus principal.

Par exemple, dans la plateforme Moodle, chaque étudiant peut être considéré comme un cas suivant un parcours d'apprentissage. La série d'événements associés à un cas spécifique est appelée une trace. Chaque ligne du tableau 1 représente un événement exécuté, comprenant des détails tels que l'identifiant de l'événement (*CaseId*), le *Timestamp* (jour, heure, minute et seconde), l'*Activité* concernée, ainsi que d'éventuels attributs supplémentaires pertinents pour l'événement selon les cas d'étude. Formellement, un journal d'événements $L = t_1, t_2, ..., t_k$ est un ensemble de $k$ traces où chaque trace $t_i$ ($1 \leq i \leq k$) est un ensemble de $n_i$ événements consécutifs $t_i =< e_{i1}, e_{i2}, ...e_{in_i} >$ réalisés par le même *CaseId*.

TABLEAU 1 – Exemple de journaux d'événements.

| ***CaseId*** | ***Timestamp*** | **Activité** |
|---|---|---|
| 1 | 2018-01-12T10:34:25 | *Course viewed* |
| 2 | 2018-01-12T10:36:25 | *Course viewed* |
| 1 | 2018-01-12T10:34:26 | *Course module viewed* |
| 1 | 2016-01-12T10:34:28 | *Submission viewed* |
| 3 | 2018-01-12T10:36:26 | *Course viewed* |
| 3 | 2018-01-12T10:36:27 | *Submission form viewed* |

De nombreuses méthodes de découverte de processus ont été proposées dans la littérature dans le but de générer automatiquement des modèles de processus. Les algorithmes de découverte de processus visent à extraire des modèles de processus à partir des logs. Ces algorithmes ont pour objectif de représenter l'ensemble des activités capturées dans les logs. Divers modèles peuvent être générés à cette fin, notamment les réseaux de Petri et les Fuzzy modèles (Aalst, 2016).

Cependant, de nombreuses études en matière de fouille de processus ont démontré que la création d'un seul modèle de processus pour un ensemble de journaux entier n'est pas idéale, notamment pour les ensembles de données très volumineux contenant des processus non structurés. Un processus non structuré est généralement piloté par un utilisateur plutôt que par un logiciel. Il en résulte de nombreux chemins possibles, mais seuls quelques-uns sont pertinents. Les techniques de *process mining* conduisent souvent à des modèles complexes et/ou surajustés, tels que le modèle en « spaghettis » ou le modèle en « fleur » identifiés dans la littérature (Aalst, 2016). Pour surmonter ces problèmes, les travaux existants ont proposé des méthodes de regroupement des traces avant la modélisation (Diamantini *et al*., 2016). La littérature propose de nombreuses approches de regroupement des traces, qui peuvent être catégorisées en trois types de techniques de regroupement en fonction de la façon dont les traces sont présentées avant le regroupement (Song *et al*., 2008 ; Zandkarimi *et al*., 2020). De plus, il existe une catégorie de regroupement dite hybride, qui intègre diverses techniques issues des méthodes mentionnées précédemment (De Koninck, De Weerdt, 2019).

La première catégorie, appelée *Trace based clustering*, regroupe les traces en fonction de leur similarité syntaxique, comme expliqué dans (Bose, Aalst, 2009a) et (Chatain *et al*., 2017). Cette approche s'inspire de la distance de Levenshtein, qui mesure la dissimilarité entre deux chaînes de caractères. Dans ce contexte, une trace peut être transformée en une autre par le biais d'opérations d'édition telles que la substitution, l'ajout ou la suppression d'événements. La distance d'édition entre deux traces est ensuite calculée comme le nombre minimum d'opérations d'édition nécessaires pour convertir une trace en une autre. Une distance d'édition plus faible indique un niveau de similarité plus élevé entre les traces. Ensuite, des algorithmes de regroupement basés sur la distance sont appliqués pour regrouper les traces en clusters distincts. Dans le domaine de l'éducation, à la fois (Laksitowening *et al*., 2023) et (Zhang *et al*., 2022) se concentrent sur les logs des étudiants pour capturer différentes caractéristiques et schémas d'apprentissage. Ils utilisent tous deux le regroupement hiérarchique comme algorithme de regroupement pour regrouper les traces des étudiants.

La deuxième catégorie est le *Model-based clustering*. Elle met l'accent directement sur la qualité des modèles découverts et la distribution des traces parmi les groupes de traces. Elle suppose que les modèles de processus précis et pertinents sont découverts à partir de sous-logs homogènes (Cadez *et al*., 2003 ; Ferreira *et al*., 2007). Le modèle de processus est considéré comme une entrée pour le regroupement afin de structurer les traces. Ces traces sont ensuite utilisées pour extraire de nouveaux modèles de processus. Les groupes de traces obtenus dépendent fortement des résultats de mesures d'évaluation de la qualité des modèles découverts (De Weerdt *et al*., 2013).

La troisième catégorie, appelée *Feature-based clustering*, implique la conversion de chaque trace en un vecteur de caractéristiques basé sur des caractéristiques prédéfinies. La similarité entre deux traces est ensuite déterminée par la similarité entre leurs vecteurs respectifs. Les méthodes existantes dans cette catégorie reposent souvent sur des métriques telles que la fréquence des événements ou la fréquence des relations de

succession directe entre les événements pour transformer les traces en vecteurs (Song *et al*., 2008). Par exemple, (Song *et al*., 2008) a analysé des traces provenant de systèmes d'information de santé, les convertissant en caractéristiques telles que le nombre d'occurrences individuelles d'événements ou le nombre de paires d'événements en succession immédiate. (Bose, Aalst, 2009b) a utilisé une technique similaire sur de plus longues sous-parties de traces, évaluant l'occurrence de motifs plus complexes tels que les répétitions, définies comme des *n-grammes* observés à différents points de la trace. Par la suite, des algorithmes de regroupement basés sur la distance sont appliqués pour regrouper les traces en clusters distincts (Zandkarimi *et al*., 2020).

Dans cet article, notre travail s'inscrit dans l'approche basée sur les caractéristiques (*Feature-based clustering*), où nous améliorons une méthode appelée *FSS Encoding* (Trabelsi *et al*., 2021). Cette méthode, initialement conçue pour les bibliothèques numériques, vise à extraire des caractéristiques des séquences en identifiant les sous-séquences fréquentes et en les encodant. L'encodage des sous-séquences fréquentes prend en compte divers paramètres pour différencier efficacement les séquences des utilisateurs des bibliothèques numériques. Un algorithme de regroupement est ensuite appliqué sur les traces converties afin d'assigner chaque trace d'apprenant au *cluster* approprié. Dans le contexte des bibliothèques numériques, la méthode *FSS encoding* a montré son efficacité pour modéliser les trajectoires des utilisateurs. Ces résultats ont été validés sur des données réelles issues de la Bibliothèque Nationale de France.

## 3. *FSS-encoding* appliqué aux traces d'apprenants utilisant Moodle

Comme mentionné dans la section 1, nous avons amélioré la méthode *FSS encoding* (désignée dorénavant comme la méthode de référence) proposée par (Trabelsi *et al*., 2021). Cette amélioration préserve davantage d'informations sur les traces ainsi que l'unicité de chaque trace, ce qui permet de découvrir de meilleurs motifs à partir des traces.

### *3.1. FSS Encoding : référence*

Nous allons décrire brièvement la méthode de référence et l'algorithme *FSS encoding* proposé dans (Trabelsi *et al*., 2021). La stratégie fondamentale de cette méthode repose sur l'hypothèse qu'une trace est caractérisée par sa ou ses sous-séquences fréquentes (FSS) la ou les plus significatives (Lu *et al*., 2019). Cette stratégie implique de regrouper les traces en fonction des sous-séquences fréquentes. Une FSS, désignée comme $< e_1, ..., e_n >$, comprend un ensemble fini d'événements de longueur $n$ ($n > 1$), où les événements sont exécutés dans l'ordre au moins deux fois.

Les traces sont converties à l'aide d'un encodage spécifique. Dans cet encodage, chaque FSS identifiée dans une trace est remplacée par son encodage correspondant. Les événements qui n'appartiennent à aucune FSS sont considérés comme non pertinents, et seules leurs positions contribuent au regroupement. Par conséquent, de tels événements dans les traces sont remplacés par la valeur 1.

La méthode de référence elle-même vise à distinguer efficacement les traces au sein de différents clusters en considérant des facteurs tels que [*1- la longueur*] et [*2- la fréquence*] des FSS, [*3- la fréquence des événements*] au sein des FSS, et [*4- la fréquence des relations de succession directe entre les événements*] dans les FSS. Cette stratégie d'encodage améliore la représentation vectorielle des traces, en mettant l'accent sur l'importance des FSS plus longues, des fréquences plus élevées, ainsi que sur l'occurrence d'événements et de relations spécifiques au sein des séquences.

Toutes les FSS extraites sont encodées dans chaque trace comme suit :

$$Encoding(FSS) = \frac{1}{f_{FSS} \sum_{i=1}^{n-1} f_{e_i} f_{e_{i+1}} f_{r_{i,i+1}}} \tag{1}$$

Où, $f_{FSS}$ est la fréquence de la FSS extraite, $n$ est sa longueur (nombre d'événements), $f_{e_i}$ est la fréquence de l'événement $e_i$ dans les logs et $f_{r_{i,i+1}}$ est la fréquence de la relation directe entre tous les événements consécutifs de la FSS dans les logs. La valeur d'encodage résultante est comprise entre $0$ et $1$. Une valeur proche de $0$ indique l'importance de la FSS dans l'ensemble des journaux d'événements.

Cette méthode d'encodage permet de distinguer les traces qui partagent la même FSS mais pas dans la même position. De plus, en remplaçant tous les événements ne faisant pas partie d'une FSS par $1$, l'information sur la position de la FSS est conservée, tout comme les écarts entre différentes FSS et la taille de la trace. Après l'encodage FSS, toutes les traces dans les logs sont converties en vecteurs. Ces vecteurs sont ensuite regroupés en fonction de leur similitude.

### *3.2. FSS Encoding : Amélioration*

L'approche de référence remplace tous les événements qui ne font pas partie d'une sous-séquence fréquente par $1$. Cela entraîne une perte d'informations concernant l'unicité des activités individuelles qui ne participent pas à un motif. Ces activités peuvent avoir une importance même si elles se produisent moins fréquemment. La simplification de la méthode de référence peut négliger la diversité et l'importance de telles activités singulières, ce qui pourrait avoir un impact sur l'exhaustivité de l'analyse. Par exemple, deux traces, $t_1 = < e_1, e_2, e_3, e_4, e_5 >$ et $t_2 = < e_0, e_1, e_2, e_3 >$, seront converties en vecteurs $[E_{FSS_1}, 1, 1]$ et $[1, E_{FSS_1}]$, respectivement, où $E_{FSS_1}$ représente l'encodage de $< e_1, e_2, e_3 >$. Les événements $e_0$, $e_4$ et $e_5$ ne seront pas pris en compte dans le regroupement.

D'autre part, notre amélioration prend en compte la fréquence et les relations à la fois des FSS et des événements individuels, offrant une représentation plus nuancée qui préserve les caractéristiques distinctives de chaque activité. Par exemple dans Moodle, un apprenant est plus susceptible de consulter un cours plusieurs fois avant de passer un quiz une seule fois. Ainsi, il est important de préserver la position et l'identité de ces activités moins fréquentes car elles pourraient contenir des informations significatives pour comprendre le processus d'apprentissage entrepris. Guidés par cela,

en utilisant notre approche améliorée, les traces $t_1$ et $t_2$ dans l'exemple précédent sont respectivement converties en vecteurs $[E_{FSS_1}, f(e_4), f(e_5)]$ et $[f(e_0), E_{FSS_1}]$. $f(a)$ représente la fréquence d'occurrence de l'activité $a$ dans toutes les traces, préservant ainsi son identité en fonction de sa fréquence qui reflète finalement sa signification.

L'algorithme 1 présente la version améliorée de la méthode de référence, décrivant l'approche proposée dans le présent article. La première étape consiste à transformer les logs originaux $R$ (voir le tableau 1) en un ensemble de traces $L$. Cet ensemble organise la séquence d'événements chronologiquement en se basant sur l'identifiant unique *CaseId*. Par exemple, en se basant sur le tableau 1, la trace correspondante de *CaseId* 1 est $<$ *Course viewed*, *Course module viewed*, *Submission viewed* $>$.

**Data :** Original log file $R$, Minimum pattern support percentage $minSup$, Minimum pattern Length $minLen$, Number of Clusters $n\_clusters$
**Result :** Log Files $F$ corresponding to resulting clusters
**begin**
- Convert $R$ to a set of traces $L$;
- From $L$, extract frequent sub-sequences $FSS$ with length $>= minLen$ and minimum support $>= minSup$;
- From $FSS$, remove $x \in FSS$ if $x$ does not exist as is in $L$;
- For $x \in FSS$, compute $Encoding(x)$ ;
- Sort $FSS$ in descending order of pattern lengths ;
- For each trace in $L$, replace any existing $FSS$ by their encoding;
- Remove traces from $L$ where no $FSS$ is found;
- For remaining traces in $L$, replace remaining activities with their frequency ;
- Scale the values in the traces using MinMax Scaler, to have a range of $[0, 1]$ ;
- Add padding of $-1$ for the traces to have same lengths ;
- Cluster the traces in $L$ into $n\_clusters$;
- Generate Log Files $F$ for resulting clusters;
- Return $F$;

**end**

**Algorithme 1 :** Algorithme d'encodage FSS amélioré

Ensuite, nous utilisons l'algorithme *PrefixSpan*pour extraire les motifs séquentiels $FSS$ à partir des logs modifiés $L$. Nous choisissons cet algorithme pour sa capacité à identifier de manière efficace les motifs fréquents dans les traces, qu'il s'agisse de motifs avec des événements contigus ou non contigus, ce qui facilite la découverte de séquences d'activités significatives.

Dans la méthode de référence, les critères de sélection des sous-séquences fréquentes avec *PrefixSpan* reposent sur les $k$-motifs les plus fréquents extraits. Cependant, cette approche entraîne une perte de précision sur la pertinence des motifs en fonction du nombre de traces qui les contiennent. Il peut en résulter des situations où tous les motifs les plus fréquents ont des pourcentages d'apparition supérieurs à 90%, ou au contraire, certains motifs ont des pourcentages d'apparition inférieurs à 50%. En outre, l'utilisation des motifs les plus fréquents peut être coûteuse en termes de calcul, car elle implique la génération et la vérification de nombreux motifs potentiels. Notre approche améliore cette méthode en affinant les critères de sélection des motifs, en intégrant deux seuils : (i) un pourcentage d'apparition minimum ($minSup$) et (ii) une longueur de motif ($minLen$). Ainsi, lors de cet affinage, l'ensemble de données est

parcouru une seule fois pour déterminer l'apparition des motifs candidats, puis seuls ceux dépassant le seuil spécifié sont conservés. L'apparition d'un motif $X$ est défini comme le rapport des traces dans lesquelles $X$ apparaît par rapport au nombre total de traces, tandis que la longueur d'un motif correspond au nombre d'activités qu'il contient. De plus, étant donné que *PrefixSpan* peut extraire des motifs qui ne sont pas présents en tant que tels dans les traces, ces motifs sont filtrés lors de la sélection.

Ensuite, l'encodage de chaque FSS est calculé en utilisant l'équation 1 et les FSS sont triées par ordre décroissant de leurs longueurs (plus le motif est long, plus il est important). Cela définit la priorité de remplacement dans l'étape suivante. Pour chaque trace, où une FSS est découverte, elle est remplacée par son encodage. Si deux FSS sont trouvées dans la même trace, la FSS la plus longue est d'abord remplacée, puis l'autre FSS est cherchée dans le reste de la trace.

Après l'encodage, les traces sans FSS trouvées sont supprimées car elles sont considérées comme non représentatives. Pour les traces restantes, les activités individuelles qui ne sont pas associées à un motif sont remplacées par leur fréquence comme expliqué précédemment. Les valeurs encodées sont normalisées entre $[0, 1]$ puis l'algorithme de clustering est exécuté sur les traces restantes, converties en vecteurs numériques. Enfin, les fichiers de logs correspondant à chaque cluster (identifiés par *CaseIds*) sont générés et renvoyés en sortie. Ces fichiers sont ensuite utilisés pour découvrir un modèle de processus décrivant le comportement général des apprenants dans chaque cluster.

En résumé, notre méthode améliore l'encodage des traces de deux manières principales. Tout d'abord, elle affine les critères de sélection des motifs en remplaçant l'approche des $k$-motifs les plus fréquents par une combinaison du pourcentage d'apparition minimum et de la longueur des sous séquences fréquentes. Ensuite, lors de la conversion des traces, les activités individuelles au sein des traces possédant des FSS sont remplacées par leurs fréquences respectives au lieu d'une valeur uniforme de 1. Cette modification préserve l'identité et la signification positionnelle des événements moins fréquents. Ces améliorations ont un impact direct sur la représentation des traces, ce qui améliore la qualité du regroupement des traces. Par conséquent, cette amélioration facilite une compréhension et une analyse plus détaillées des scénarios d'apprentissage au sein de chaque cluster.

## 4. Prétraitement des données

Étant donné que les données avec lesquelles nous travaillons concernent les logs issus de la plate-forme Moodle. Cette section est ainsi dédiée à une explication détaillée des étapes de collecte et de traitement de ces données.

### *4.1. Collecte et prétraitement des données*

Moodle est une plate-forme d'apprentissage en ligne bien connu et largement utilisé dans les universités et les établissements éducatifs. Elle contient un système de journalisation qui capture toutes les interactions des utilisateurs avec le système. Dans

le cadre de nos travaux, les journaux d'événements Moodle de 471 étudiants, inscrits à des cours du département d'informatique et de génie informatique de l'Université Frederick à Chypre de 2018 à 2022, ont été collectés. Les logs collectés ont été nettoyés pour ne conserver que les actions effectuées par les étudiants sur Moodle pendant leurs études, telles que suivre des cours, passer des tests et effectuer des devoirs. En effet, les logs initiaux contenaient les actions effectuées par tous les utilisateurs du système (étudiants, instructeurs, assistants, gestionnaires, etc.). De plus, un identifiant unique a été créé pour chaque étudiant car l'identification des étudiants change en fonction des années dans les logs initiaux.

La structure d'un fichier logs est présentée dans le tableau 2. Le *Regnum* est le numéro d'inscription, utilisé comme identifiant unique pour suivre le parcours d'un étudiant à travers différents cours et différentes années, c'est-à-dire qu'il est utilisé comme *CaseId*. Le *Timestamp* enregistre l'heure exacte de chaque événement effectué par les étudiants. Il est utilisé pour ordonner les événements. Le « Nom de l'événement » est utilisé comme Activité et le « Contexte de l'événement » donne des informations sur la ressource d'apprentissage concernée (fichier, devoir, dossier, etc.) par l'événement. Enfin, la colonne « Description » décrit l'événement de manière plus détaillée.

TABLEAU 2 – Structure d'une ligne d'un fichier log

| *Regnum* | *Timestamp* | *Event Context* | *Event Name* | *Description* |
|---|---|---|---|---|

Le nombre initial d'activités était de 65, comprenant des événements liés aux actions de cours, à la réalisation de quiz, à la soumission de devoirs, aux discussions et chats, à la consultation de profils, et autres. Seuls 14 événements sont conservés, comme indiqué dans le tableau 3. Ces événements ont été choisis car ils sont représentatifs des actions telles que l'achèvement d'un devoir ou d'une ressource d'apprentissage, l'évaluation, la réception de feedback, l'étude et l'exploration. Le journal est filtré en fonction des événements choisis pour finalement aboutir à 471 étudiants avec un total de 3942 traces pour la période de 2018 à 2022.

TABLEAU 3 – Activités retenues dans les logs Moodle.

| **Noms des activités** | |
|---|---|
| *A submission has been submitted* | *Quiz attempt submitted* |
| *Course activity completion updated* | *Course module viewed* |
| *Zip archive of folder downloaded* | *Content page viewed* |
| *Clicked join meeting button* | *Course summary viewed* |
| *Course module instance list viewed* | *Sessions viewed* |
| *Lesson started* | *Lesson resumed* |
| *Feedback viewed* | *Course viewed* |

### *4.2. Enrichissement sémantique des données*

Étant donné que les logs collectés observent l'interaction entre les étudiants et la plate-forme Moodle dans des différents cours, une étape d'enrichissement sémantique a été réalisée pour définir de nouvelles activités (au sens actions observées par la fouille de processus) qui donne plus d'informations sur le sens de l'action pédagogique effectuée. Nous appelons cette activité « activité sémantique ». La création de ces activités est basée sur des règles en prenant en compte le « contexte de l'événement », le « nom de l'événement » et la « description » des logs d'origine. Les détails de cette étape de transformation sont en dehors du cadre de cet article. Les activités peuvent avoir l'une des 12 valeurs présentées dans le tableau 4. Nous différencions deux types d'activités : passive et active (indiquées respectivement par _P et _A). Lorsqu'un étudiant télécharge du matériel de cours pour l'étudier, cette action est qualifiée de « passive » car elle ne garantit pas nécessairement que l'étudiant va ensuite consulter la ressource téléchargée. En revanche, lorsque qu'un étudiant soumet un devoir, nous présumons qu'il a terminé les exercices, ce qui constitue une action « active ».

TABLEAU 4 – Les activités sémantiques générées

| ***Les activités sémantiques*** | | | |
|---|---|---|---|
| *Study_P* | *Study_A* | *Revise* | *Expand* |
| *Exercise_P* | *Exercise_A* | *View* | *Interact* |
| *Assess_P* | *Assess_A* | *Feedback* | *Apply* |

### *4.3. Préparation des Traces*

Cette section décrit l'étape initiale de l'algorithme 1. Pour effectuer le regroupement, les traces doivent être extraites des logs. Comme chaque log correspond aux activités ou événements réalisées dans un cours, nous définissons une trace comme une séquence ordonnée d'événements réalisés par un étudiant dans un cours. Ainsi, le fichier de traces utilisé pour effectuer le regroupement est structuré comme illustré dans le tableau 5, où *CaseId* est la valeur unique d'un étudiant se comportant dans un cours et la trace $t_i$ ($1 \leq i \leq k$) est une séquence ordonnée de $n_i$ événements (transformé en « activité sémantique ») $t_i =< se_{i1}, se_{i2}, ..., se_{in_i} >$ réalisés par le même *CaseId*.

TABLEAU 5 – La structure d'un fichier de trace.

| *CaseId* | *Trace* |
|---|---|
| *1_course1* | <*View*, *Exercise_P*, *Assess_A* > |
| *1_course2* | <*View*, *View*, *Study_P*, *Study_A* > |
| *2_course1* | <*Interact* > |

## 5. Implémentation et résultats

Dans ce qui suit, nous évaluons les améliorations de notre approche au niveau de l'extraction et l'enodage des motifs ainsi qu'au niveau des clusters et des modèles découverts.

### *5.1. Extraction et encodage des sous séquences fréquentes*

Le tableau 6 compare la méthode de référence et notre approche *FSS encoding* améliorée au niveau de l'extraction des motifs et de l'encodage des traces. Nos critères de sélection affinés offrent un meilleur contrôle sur l'apparition minimum et la longueur des motifs, ce qui conduit à plus de motifs et à des motifs plus longs avec une apparition élevée. Les traces encodées dans notre approche sont plus courtes, mettant en avant des motifs plus longs et plus significatifs, surpassant la méthode de base à la fois dans l'extraction et l'encodage des motifs, comme le reflètent les résultats ultérieurs du regroupement des traces.

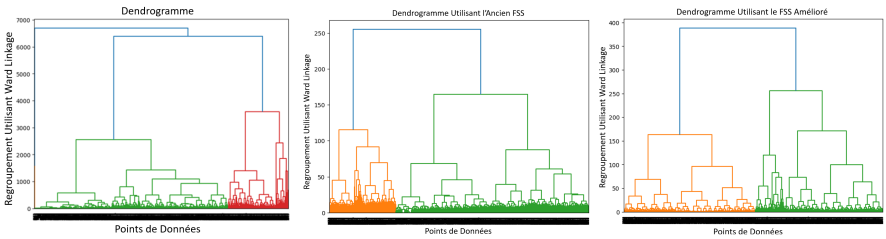
TABLEAU 6 – Comparaison des motifs extraits.

| | **Référence** | **FSS amélioré** |
|---|---|---|
| Paramètres | K = 100 | (MinSup = 80%, MinLen = 2) |
| Nb de motifs extraits | 100 | 1412 |
| Nb de motifs existants dans les traces | 32 | 248 |
| *Parmi les motifs qui existent tels quels dans les traces* | | |
| [Min - Max] Longueur du motif | [1 - 6] | [2 - 9] |
| [Min - Max] Apparition du motif % | [86% - 100%] | [80% - 95%] |
| *Niveau trace* | | |
| [Min - Max] Longueur originale de la trace | [2 - 5599] | [2 - 5599] |
| [Min - Max] Longueur des traces encodées | [1 - 2484] | [1 - 1618] |
| Nb de traces sans FSS | 49 | 45 |

### *5.2. Regroupement des Traces*

Dans la dernière étape de l'algorithme 1, les traces encodées à l'aide la méthode *FSS-encoding* sont regroupées à l'aide de l'algorithme de clustering *Hierarchical Agglomerative Clustering* (HAC) avec l'option « *ward linkage* », connue pour fusionner des clusters similaires selon une approche ascendante. Nous démontrons l'efficacité de notre méthode grâce à une comparaison avec la méthode de référence et une autre basée sur la fréquence d'activité (Song *et al.*, 2008). Cette dernière transforme les traces en vecteurs binaires. La longueur du vecteur est égale au nombre d'activités uniques, fournissant une représentation binaire de la présence de l'activité dans chaque trace. En ce qui concerne le regroupement, le nombre optimal de clusters, déterminé par l'analyse du dendrogramme, est de 3 pour toutes les approches, comme le montre la figure 2. Les métriques d'évaluation, y compris le coefficient de silhouette ($-1$ à 1, le

plus élevé étant le meilleur) et l'indice de Davies Bouldin (plus bas étant le meilleur), révèlent la qualité des clusters résultants, comme présenté dans le tableau 7.




(a) Fréquence des activités (b) FSS référence (c) FSS amélioré

FIGURE 2 – Les dendrogrammes générés par la méthode HAC en utilisant le *ward linkage*

TABLEAU 7 – La qualité du clustering de chaque méthode et les mesures de Silhouette pour les différents clusters

| | | Fréquence des activités | FSS-référence | FSS-amélioré |
|---|---|---|---|---|
| Nb Finale de Traces | | 3942 | 3893 | 3897 |
| Silhouette Coefficient | | 0.546 | 0.117 | 0.360 |
| Davies Bouldin Index | | 0.720 | 2.43 | 1.15 |
| *Détail de l'analyse Silhouette* | | | | |
| Cluster 0 | Nb de Traces | 710 | 1000 | 1496 |
| | Silhouette | 0.156 | -0.065 | 0.260 |
| Cluster 1 | Nb de Traces | 3 | 1482 | 1961 |
| | Silhouette | 0.520 | 0.038 | 0.470 |
| Cluster 2 | Nb de Traces | 3229 | 1411 | 440 |
| | Silhouette | 0.632 | 0.408 | 0.250 |

Bien que l'on puisse initialement penser que le coefficient de silhouette est meilleur sans l'encodage des FSS, une analyse approfondie du tableau 7 révèle des interprétations potentiellement erronées des valeurs numériques. La méthode basée sur la fréquence des activités regroupe presque tous les éléments dans un seul cluster de traces, rendant ses résultats peu pertinents. En revanche, l'approche de référence divise les traces en différents clusters, mais ceux-ci manquent de séparation claire, comme en témoigne leur valeur de silhouette tandis que notre méthode combine des clusters bien séparés avec des coefficients de silhouette acceptables pour chaque cluster.

### *5.3. Découverte des modèles de comportements des apprenants*

Nous utilisons l'algorithme *Fuzzy Miner*, implémenté dans l'outil ProM (Aalst, 2016) pour découvrir des modèles de processus de chaque cluster. Nous avons choisi l'algorithme de découverte *Fuzzy Miner* pour sa capacité à générer des modèles simplifiés mettant l'accent sur les nœuds significatifs et les arcs bien corrélées (Aalst,

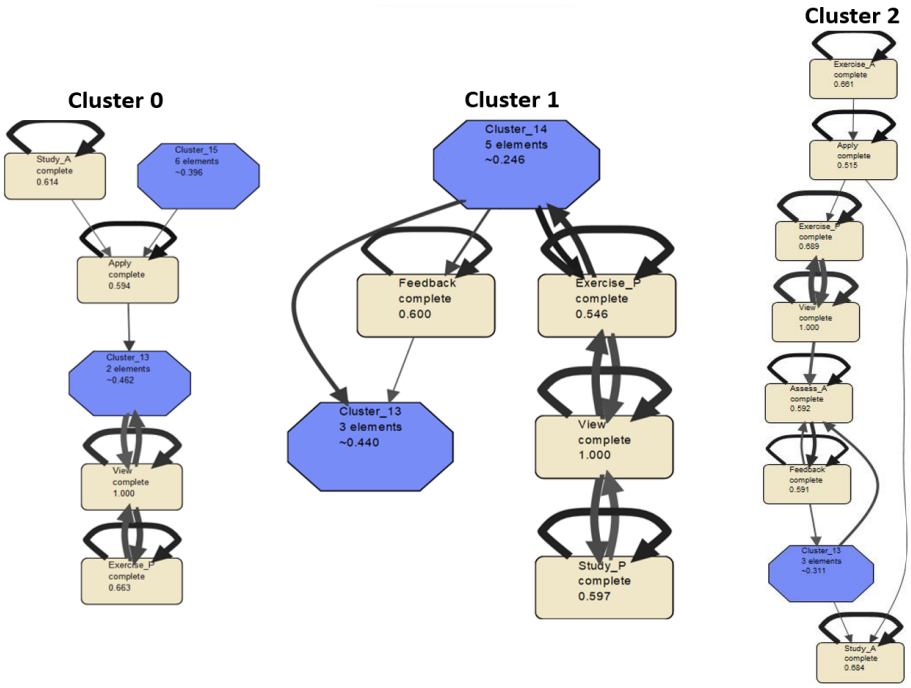



FIGURE 3 – Les modèles *Fuzzy* découverts à partir des clusters.

2016). Avec *Fuzzy Miner*, les nœuds les moins significatifs (moins fréquentes) mais fortement corrélés sont agrégés, c'est-à-dire cachés dans des clusters ayant la couleur violet au sein du modèle simplifié. La figure 3 affiche les modèles des clusters 0, 1 et 2, simplifiés en utilisant une métrique signification des nœuds élevée, ce qui conduit à l'agrégation de certains nœuds.

L'interprétation qualitative des modèles nous montre que les apprenants du cluster 1, représentant la majorité des apprenants, s'engagent principalement dans des activités routinières telles que la consultation, l'étude et la résolution d'exercices. En revanche, les étudiants du cluster 0, le deuxième plus grand cluster, regroupe des apprenants qui non seulement effectuent ces tâches routinières, mais présentent également un profil distinct en appliquant activement leurs connaissances, souvent à travers des soumissions de projets. Enfin, le cluster 2, avec le plus petit nombre d'apprenants, est composé d'individus qui préfèrent les quiz et les tests dans le cadre de leurs parcours d'apprentissage.

## 6. Conclusion et perspectives

Cette étude s'inscrit dans le prolongement d'un travail précédent présentant une plateforme visant à fournir des parcours d'apprentissage adaptatifs personnalisés, prenant en compte l'objectif de l'apprenant et exploitant l'expérience d'apprentissage

des apprenants précédents. En utilisant la fouille de processus, nous extrayons les parcours d'apprentissage passées grâce à la découverte de scénarios d'apprentissage. Cependant, traiter les données Moodle non structurées et volumineuses, qui possèdent des caractéristiques d'apprentissage spécifiques, constitue un défi, rendant le regroupement des traces crucial. Ainsi, notre approche améliore une méthode de regroupement des traces basée sur les sous-séquences fréquentes (FSS) en raffinant la sélection des motifs, en préservant notamment l'unicité des événements moins fréquents. Appliquée aux journaux Moodle, notre méthode montre des améliorations significatives, générant plus de motifs et des motifs plus longs, influençant les résultats d'encodage et conduisant à de meilleurs clusters comme le traduit le coefficient de silhouette.

Les clusters identifiés révèlent trois scénarios d'apprentissage distincts : l'un caractérisé par une concentration sur l'étude et la résolution d'exercices, un autre par l'application des connaissances acquises à travers des projets, et un troisième par une préférence pour réaliser plus d'évaluations. Ces scénarios fournissent des informations précieuses pour produire des recommandations personnalisées. Une évaluation par des experts des clusters identifiés, accompagnée d'une analyse approfondie des scénarios d'apprentissage, complétera bien notre démarche. Les perspectives de ces travaux visent à intégrer ces résultats dans le cadre de recommandation, en exploitant les expériences d'apprentissage passées pour un guidage plus efficace. Il convient de noter qu'une série de tests approfondis d'algorithmes de regroupement et de critères pour le clustering hiérarchique ont précédé la sélection de l'approche la plus performante présentée dans ce travail.

## Bibliographie

Aalst W. Van der. (2016). *Process mining: data science in action*. Berlin, Heidelberg, Springer.

Aggarwal C. C. (2016). *Recommender Systems*. Cham, Springer International Publishing. Consulté sur http://link.springer.com/10.1007/978-3-319-29659-3

Bose R. J. C., Aalst W. M. Van der. (2009a). Context aware trace clustering: Towards improving process mining results. In *Proceedings of the international conference on data mining*, p. 401–412.

Bose R. J. C., Aalst W. M. van der. (2009b). Trace clustering based on conserved patterns: Towards achieving better process models. In *International conference on business process management*, p. 170–181.

Cadez I., Heckerman D., Meek C., Smyth P., White S. (2003). Model-based clustering and visualization of navigation patterns on a web site. *Data mining and knowledge discovery*, vol. 7, n° 4, p. 399–424.

Cenka N., Anggun B. (2022, mars). Analysing student behaviour in a learning management system using a process mining approach. *Knowledge Management & E-Learning: An International Journal*, vol. 14, n° 1, p. 62–80.

Chatain T., Carmona J., Van Dongen B. (2017). Alignment-based trace clustering. In *International conference on conceptual modeling*, p. 295–308.

De Koninck P., De Weerdt J. (2019). Scalable mixed-paradigm trace clustering using super-instances. In *2019 international conference on process mining*, p. 17–24.

De Weerdt J., Vanden Broucke S., Vanthienen J., Baesens B. (2013). Active trace clustering for improved process discovery. *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, n° 12, p. 2708–2720.

Diamantini C., Genga L., Potena D. (2016). Behavioral process mining for unstructured processes. *Journal of Intelligent Information Systems*, vol. 47, n° 1, p. 5–32.

Ferreira D., Zacarias M., Malheiros M., Ferreira P. (2007). Approaching process mining with sequence clustering: Experiments and findings. In *International conference on business process management*, p. 360–374.

Joudieh N., Eteokleous N., Champagnat R., Rabah M., Nowakowski S. (2023). Employing a process mining approach to recommend personalized adaptive learning paths in blended-learning environments. In *12th international conference in open and distance learning, athens, greece.*

Laksitowening K. A., Prasetya M. D., Suwawi D. D. J., Herdiani A. *et al.* (2023). Capturing students' dynamic learning pattern based on activity logs using hierarchical clustering. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, n° 1, p. 34–40.

Lu X., Tabatabaei S. A., Hoogendoorn M., Reijers H. A. (2019). Trace clustering on very large event data in healthcare using frequent sequence patterns. In *International conference on business process management*, p. 198–215.

Song M., Günther C. W., Aalst W. M. Van der. (2008). Trace clustering in process mining. In *International conference on business process management*, p. 109–120.

Trabelsi M., Suire C., Morcos J., Champagnat R. (2021). A new methodology to bring out typical users interactions in digital libraries. In *2021 acm/ieee joint conference on digital libraries (jcdl)*, p. 11–20.

Zandkarimi F., Rehse J.-R., Soudmand P., Hoehle H. (2020). A generic framework for trace clustering in process mining. In *2020 2nd international conference on process mining*, p. 177–184.

Zhang T., Taub M., Chen Z. (2022). A multi-level trace clustering analysis scheme for measuring students' self-regulated learning behavior in a mastery-based online learning environment. In *Lak22: 12th international learning analytics and knowledge conference*, p. 197–207.

# L'effet de la complexité visuelle de l'information sur l'intention et le comportement de mobilité urbaine

**Thomas Chambon[1], Ulysse Soulat[2], Jeanne Lallement [2], Jean-Loup Guillaume [1]**

*1. Laboratoire Informatique, Image et Interaction (L3i), La Rochelle Université, 23 Avenue Albert Einstein, 17000 La Rochelle, France, thomas.chambon1@univ-lr.fr, jean-loup.guillaume@univ-lr.fr*

*2. Laboratoire Usages du Numérique pour le Développement Durable (NUDD), La Rochelle Université, 39 rue de Vaux De Foletier, 17000 La Rochelle, France, aymeric-ulysse.soulat@univ-lr.fr, jeanne.lallement@univ-lr.fr*

*Cet article est une synthèse de l'article :*
*Thomas Chambon, Ulysse Soulat, Jeanne Lallement, Jean-Loup Guillaume: The Effect of Visual Information Complexity on Urban Mobility Intention and Behavior. RCIS 2023: 452-466*

## 1. Introduction

Malgré une baisse des émissions de gaz à effet de serre provenant des transports, le rapport du GIEC de 2022 prévient que le réchauffement climatique en Europe augmentera plus que la moyenne mondiale. Le rapport insiste également sur l'efficacité de l'adoption de choix de mobilité urbaine douce. Malgré les efforts croissants des organisations et de la recherche pour transformer les intentions comportementales en comportements vertueux, il reste encore beaucoup à découvrir pour combler ce "Green gap" qui appelle à un travail des chercheurs en comportement durable (Trudel, 2019). Notre recherche s'intéresse particulièrement à l'effet de l'adoption d'une application de self-tracking de son empreinte carbone et aux conséquences de la complexité de l'information CO2 sur le comportement des usagers. Nous combinons des méthodes issues de l'informatique et des sciences de gestion pour examiner les processus décisionnels et comportementaux. L'association des deux disciplines offre une perspective originale d'examen des effets de la complexité visuelle d'une application.

## 2. Méthodologie et résultats

Notre étude mobilise deux champs théoriques complémentaires, relatifs à la perception de la complexité visuelle et l'adoption d'une technologie en lien avec cette complexité. D'une part, la complexité visuelle est définie par la quantité d'éléments ainsi que leur niveau de détail sur l'écran de l'application. Elle peut être également définie comme la diversité des éléments visuels composant un stimulus. Les caractéristiques de l'information (i.e., la couleur, l'orientation, la fréquence spatiale, la luminance et le mouvement) contribuent à sa perception et influence l'attention visuelle. La plupart des recherches sur le sujet, mobilisant la théorie de la complexité visuelle (Berlyne, 1960), indique qu'un stimulus d'une complexité visuelle modérée (par opposition à simple ou complexe) est plus efficace sur les attitudes et les intentions d'achat. D'autre part, l'adoption des applications de self-tracking peut être étudiée sous l'angle de l'acceptation d'une technologie (Venkatesh, 2012). Nous examinons successivement les effets de l'acceptation d'une technologie de self-tracking selon sa complexité visuelle et les conséquences sur le comportement.

Dans une première expérimentation auprès de 362 individus, nous avons observé qu'une complexité visuelle modérée a un effet plus significatif sur l'intention d'adopter un comportement responsable en matière de transport que des pages d'accueil d'une complexité visuelle complexe ou simple. Dans la seconde expérimentation auprès de 51 utilisateurs réguliers de l'application, nous avons cependant constaté qu'une complexité visuelle modérée n'a pas d'effet plus significatif sur le comportement de mobilité responsable que des pages d'accueil de complexité visuelle plus élevée ou plus faible. Au regard de ces résultats nous souhaiterions intégrer des fonctionnalités avec des composants tels que la comparaison sociale et l'individualisation de l'information afin d'augmenter le taux d'utilisation de l'application.

### Références

Berlyne D. E. (1960). *Conflict, arousal, and curiosity.* McGraw-Hill Book Company,

Trudel R. (2019). Sustainable consumer behavior. *Consumer Psychology Review,* 2 (1):85–96.

Venkatesh., Y. L. Thong, and Xin Xu. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly,* 36(1):157–178.

# Stratégies open-sources : opportunités et limitations dans le domaine des *Large Language Models* (LLM)

**Robert Viseur**[1]

*1. Service TIC, FWEG, UMONS*
*17 place Warocqué, B-7000 Mons, Belgique*
*robert.viseur@umons.ac.be*

*RÉSUMÉ.*
*L'avènement des modèles linguistiques de grande envergure (LLM), tels que GPT d'OpenAI, a marqué une avancée significative dans le domaine de l'intelligence artificielle. Ces développements ont été accompagnés par une réflexion sur l'importance des stratégies open-sources dans la recherche et le développement des LLM. Notre étude explore cette dynamique, mettant en lumière les bénéfices et les défis associés à l'adoption d'approches open-sources dans la création et l'utilisation de LLM. Nous examinons les différentes manières dont les données, les modèles et les applications sont partagées et développées de manière ouverte, contribuant à l'innovation et à l'amélioration continue dans le secteur. En dépit de la tendance à la privatisation et à la fermeture des modèles, ce papier argumente en faveur du potentiel des stratégies open-sources pour favoriser une intelligence artificielle éthique, transparente et accessible. En analysant les pratiques actuelles et en proposant une réflexion sur l'avenir de l'open-source dans le développement des LLM, nous soulignons comment ces stratégies peuvent répondre à divers besoins utilisateurs, de la personnalisation à la réduction des coûts, tout en stimulant l'innovation collaborative.*

*ABSTRACT.*

*The advent of Large Language Models (LLMs) such as OpenAI's GPT has marked a significant advancement in the field of artificial intelligence. These developments have been accompanied by a reflection on the importance of open-source strategies in the research and development of LLMs. Our study explores this dynamic, highlighting the benefits and challenges associated with adopting open-source approaches in the creation and use of LLMs. We examine the various ways in which data, models, and applications are shared and developed openly, contributing to innovation and continuous improvement in the sector. Despite the trend towards the privatisation and closing of models, this paper argues in favour of the potential of open-source strategies to foster an ethical, transparent, and accessible artificial intelligence. By analysing current practices and offering reflections on the future of open-source in the development of LLMs, we underline how these strategies can meet diverse user needs, from personalisation to cost reduction, whilst stimulating collaborative innovation.*

*Mots-clés : intelligence artificielle, LLM, FLOSS, éthique.*

*KEYWORDS: artificial intelligence, LLM, FLOSS, ethics.*

## 1. Introduction

L'intelligence artificielle peut être vue comme « *un artefact informatique construit grâce à l'intervention humaine, qui pense ou agit comme les humains, ou comme nous nous attendons à ce que les humains pensent ou agissent* » (Dignum, 2019). Parmi les courants qui la traversent, l'apprentissage logiciel, ou « *machine learning* », a connu une popularité croissante suite à l'essor des réseaux de neurones profonds, ou « *deep learning* », et de leurs applications. Parmi celles-ci, citons les algorithmes de type Transformers, dont sont issus les *Large Language Models* (LLM). Parmi les LLM, le modèle GPT, développé par OpenAI, utilisé notamment dans l'agent conversationnel ChatGPT, est sans doute le plus populaire aujourd'hui. D'autres entreprises sont venues par la suite concurrencer OpenAI : Google (Bard, Gemini), Anthropic (Claude), Mistral (Mistral, Le Chat), META (Llama)… Parmi ces propositions, certaines se distinguent par leur caractère open-source. Cependant, force est de constater la tendance actuelle à la fermeture des modèles précédemment ouverts ou libres (OpenAI, Mistral...). Cette évolution traduit-elle un légitime désintérêt pour ce type d'approche collaborative ? Ce papier propose de faire le point sur l'intérêt des stratégies open-sources dans le domaine des LLM.

Notre article est décomposé en trois sections. Dans une première section, nous proposons un état de l'art relatif aux *Large Language Models* (LLM) et aux stratégies dites open-sources. Ensuite, nous inventorions les pratiques open-sources dans le domaine des IA génératives (*text-to-text*) et en analysons l'intérêt. Enfin, dans une troisième et dernière section, nous concluons.

## 2. Revue de littérature

Cette revue de littérature se focalise sur deux notions : les *Large Language Models* (LLM) et les stratégies dites open-sources.

GPT est « *un modèle linguistique autorégressif de troisième génération qui utilise l'apprentissage profond pour produire des textes semblables à ceux des humains* » (Floridi & Chiriatti, 2020). Ce modèle linguistique est formé par entraînement « *sur un ensemble de données non étiquetées composé de textes, tels que Wikipédia et de nombreux autres sites, principalement en anglais, mais aussi dans d'autres langues* » (Floridi & Chiriatti, 2020). OpenAI exploite un ensemble diversifié de jeux de données incluant, pour GPT-3, le Common Crawl (60 % des données d'entraînement), WebText (22%), Books1/Books2 (16%) et Wikipédia (3%) (Brown et al., 2020). Le modèle produit est utilisable au travers de ChatGPT (version gratuite ou payante : ChatGPT Plus) ou au travers d'une API payante.

L'éthique des intelligences artificielles génératives (IAG) est associée à des critères d'équité (« *fairness* »), de transparence (« *transparency* ») et de responsabilité (« *accountability* ») (Ferrara, 2023). Elle porte tant sur la conception que sur l'utilisation des IA. Les IA génératives font en particulier l'objet d'efforts importants en matière de lutte contre les biais, définis comme « *la présence de déformations systématiques, d'erreurs d'attribution ou de distorsions factuelles qui favorisent certains groupes ou idées, perpétuent des stéréotypes ou font des suppositions incorrectes basées sur des schémas appris* », ce qu'une ouverture des données, des méthodologies et des outils facilite (Ferrara, 2023).

Les logiciels libres et open-sources recouvrent un modèle d'innovation construit par et pour les utilisateurs-innovateurs (Jullien et al., 2022). Plutôt que de privatiser le logiciel par le recours à la propriété intellectuelle, il s'appuie sur cette dernière pour « *organiser l'évolution continue de la demande et de l'innovation* ». S'ils sont pensés par et pour les utilisateurs, ces logiciels permettent cependant le développement de stratégies commerciales, dites stratégies open-sources. La proposition de valeur associée au modèle d'affaires recouvre des prestations incluant le développement sur mesure, l'édition logicielle, la production de modules spécialisés ou encore l'hébergement des applications (Jullien & Viseur, 2021). Ce modèle de développement a fait l'objet d'une extension, qualifiée d'« *open source innovation* », au-delà du seul logiciel (Pénin, 2011). Des stratégies open-sources sont dès lors déployables aussi pour des données (open-data) ou du matériel (open-hardware).

Jullien et Viseur (2021) analysent les modèles d'affaires open-sources sous l'angle de la segmentation des besoins des utilisateurs, et du coût total de possession de la solution (Shaikh & Cornford, 2011). L'intérêt du FLOSS (*Free Libre Open Source Software*) dépend en effet des besoins des clients, lesquels conditionnent les modèles d'affaires praticables. Quatre types de besoins sont identifiés par les auteurs : « *Contrôle* » (nécessité d'un haut degré de personnalisation impliquant du développement sur mesure), « *Lock-out* » (peu ou prou de besoins de personnalisation mais évitement du « *vendor lock-in* »), « *Lean* » (personnalisation de masse) et « *Prix* » (recherche d'un produit standardisé à prix modique). Le FLOSS est d'autant plus intéressant que le besoin de maîtrise de la solution (Contrôle, Lock-out, voire Lean) est important.

## 3. Stratégies open-sources appliquées aux LLM

Lors du développement et de l'exploitation d'une intelligence artificielle générative (IAG), les stratégies open-sources peuvent porter sur trois artefacts : les données, les modèles et les applications. Elles s'ajoutent à des stratégies d'innovation ouverte plus classiques, notamment en matière d'optimisation des infrastructures (p. ex. Open Compute Project).

Données :

Le jeu de données partagé de référence est le Common Crawl (CC). Alimenté par la Common Crawl Foundation, il est couvert par les Terms of Use de cette dernière. Il n'est pas strictement open-source puisque, d'une part, il contient des données issues du Web (donc couverte par le droit d'auteur de leurs créateurs), d'autre part, il n'est pas couvert par une licence libre. Le fichier publié fin décembre 2023 contient 3,35 milliards de documents pour un total d'environ 125 TiB après compression. Le partage des jeux de données permet dès lors d'éviter que tous les réutilisateurs doivent mettre en place une coûteuse infrastructure de collecte de données. Le robot d'exploration associé au CC, issus d'un *fork* du moteur de recherche open-source Nutch est cependant publié sous licence libre[1].

Le développement des LLM s'est accompagné de la publication de nombreux jeux de données (Liang et al., 2014). Le Common Crawl est en effet composé de données de qualité très variable. Deux stratégies sont dès lors déployées. La première stratégie consiste à filtrer le Common Crawl. C'est notamment ce qui est

[1] Cf. https://github.com/commoncrawl/cc-nutch-example.

appliqué par Google avec le C4 Colossal Clean Crawled Corpus (Dodge et al., 2021). Y sont par exemple filtrés les documents comportant des mots issus d'une liste de mots bannis. Au final, ce jeu de données favorise les sources de qualité comme les sites de presse (NYTimes, LATimes...) ou les plateformes de contenus scientifiques (PLoS, Springer...). La seconde stratégie consiste à produire de nouveaux jeux de données porteurs de qualités spécifiques. L'OpenWebText reprend ainsi l'esprit du WebText utilisé par OpenAI. Il est alimenté par des informations plébiscitées par les utilisateurs de Reddit (Liang et al., 2014).

L'ouverture des données en facilite l'audit et simplifie l'identification des biais induits (Ferrara, 2023). Les dispositifs de filtrage sont-ils proportionnées ou conduisent-ils à invisibiliser certaines communautés (Dodge et al., 2021) ? Les jeux de données incluent-ils des contenus toxiques (Liang et al., 2024) ? La publication des règles et des logiciels de filtrage permet ainsi une amélioration continue du processus, et d'aller vers des intelligences artificielles plus éthiques.

Modèles :

Plusieurs organisations proposent des LLM sous licence libre et open-source. C'est notamment le cas de Google (T5) et, jusqu'à récemment, de Mistral[2]. Ces modèles tendent à être diffusés sous des licences permissives telles que la licence Apache ou la licence MIT. Cela autorise les utilisateurs à intégrer le modèle dans leurs applications, tel quel ou après une étape de spécialisation (*fine tuning*), avec un accès à l'architecture, à la stratégie d'entraînement et aux poids. Notons l'usage de licences « *partly open* » (West, 2003) par certains de ces acteurs, à l'image de META[3] (Llama), notamment justifié par des considérations éthiques (voir par exemple la Responsible AI Licenses ; Contractor et al., 2022).

Plusieurs bénéfices peuvent être associés à ces LLM open-sources. Le premier bénéfice est un gain de visibilité pour l'entreprise qui publie ce modèle, grâce à la plus grande diffusion de la marque associée à l'éditeur open-source (Jullien & Viseur, 2021). Le second bénéfice découle de l'accélération de la diffusion du modèle, par téléchargement ou inclusion au sein de plateformes d'exécution (Amazon Sagemaker, NVIDIA NeMo…). D'une part, la disponibilité accrue de ces modèles stimule les retours des utilisateurs quant à leurs performances. Par exemple, les LLM font l'objet d'évaluations quant à leurs capacités ou leurs limitations (p. ex. hallucinations[4]). D'autre part, cette disponibilité des modèles permet d'accélérer le rythme d'innovation. Les améliorations portent par exemple sur la réduction de la taille des modèles (Eldan & Li, 2023). Le troisième bénéfice a trait au partage des coûts entre partenaires (si le modèle open-source est construit collaborativement). Les intelligences artificielles génératives suscitent en effet des inquiétudes en matière d'impact environnemental (Sundberg, 2024). Dès lors que l'entraînement des intelligences artificielles compterait actuellement pour 20 à 40 % de leur consommation (IEA, 2024), la création de modèles en consortium permettrait d'en réduire l'impact environnemental.

Applications :

Les besoins d'intégration des LLM entraîne la création de logiciels capables de les exploiter, soit sous la forme de progiciels (p. ex. agents conversationnels :

[2] Cf. https://github.com/eugeneyan/open-llms.
[3] Cf. https://opensource.org/blog/metas-llama-2-license-is-not-open-source.
[4] Cf. https://github.com/vectara/hallucination-leaderboard.

GPT4All), soit sous la forme de plateformes d'exécution multi-modèles (p. ex. Ollama, Hugging Face et LangChain). Ces logiciels suivent alors des logiques open-sources plus classiques.

## 4. Conclusion

Les besoins clients identifiés par Jullien et Viseur (2021) permettent de discuter l'intérêt des stratégies open-sources pour les IAG (cf. Tableau 1). Les clients à logique « Contrôle » sont davantage sensibles aux questions de personnalisation, de confidentialité et de sécurité. Ils ont un intérêt à pouvoir accéder à des jeux de données spécifiques, incluant des données internes à l'organisation, permettant la spécialisation de modèles génériques. Les clients à logique « Lock-Out » sont peu ou prou intéressés par la personnalisation des IA génératives. Le caractère open-source des modèles et des applications leur garantit par contre une relative indépendance vis-à-vis des prestataires informatiques. Les clients à logique « Lean » peuvent être intéressés par des modèles spécialisés, utilisés en combinaison, pour satisfaire leur exigence de personnalisation à coût maîtrisé. Par ailleurs, les modèles open-sources peuvent leur permettre de bénéficier de technologies en évolution rapide et d'optimiser leurs coûts d'usage. Les clients à logique « Prix » seront peu ou prou intéressés par le caractère open-source des modèles et applications. Ils privilégieront plutôt les API ou les applications web leur permettant un déploiement rapide (SaaS) et à moindre coût (tarification au *pay-per-use*).

*Tableau 1. Utilisation de stratégies d'ouverture de type open-source en fonction des besoins des utilisateurs (grisé : technologies closed-source).*

| | **Contrôle** | **Lock-Out** | **Lean** | **Prix** |
|---|---|---|---|---|
| **Données** | Jeux de données partagés (Common Crawl, C4...) | - | - | - |
| **Modèles** | LLM open-source (LLama 2, Mistral 7B…) avec *fine tuning* | LLM open-source (LLama 2, Mistral 7B…) | LLM (open-source ou non) spécialisés et combinés | API LLM (Mistral, GPT...) |
| **Applications** | Agents conversationnels open-source (GPT4All…)<br>Applications spécifiques (Ollama, LangChain...) | Agents conversationnels open-source (GPT4All...) | Applications spécifiques (Ollama, LangChain...) | ChatGPT (OpenAI, Microsoft Azure), Copilot |

Reste le cas des utilisateurs-innovateurs, assimilés par Jullien et Viseur (2021) à des utilisateurs de pointe au sens de von Hippel, développant de nouveaux LLM pour leurs besoins propres. Il recouvre des situations comme celle de META. Le développement de Llama est actuellement internalisé par l'entreprise. Cependant, il sera intéressant de suivre dans quelle mesure un modèle de fondation (Riehle, 2010), typique des FLOSS (Apache, Eclipse, Linux…), émergera en vue de partager les coûts de recherche et de création des LLM ou d'autres modèles d'intelligence artificielle nécessitant d'importante ressources informatiques (à l'image de PyTorch, transféré par META vers Linux Foundation).

## 5. Références


Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165. https://doi.org/10.48550/arXiv.2005.14165.

Contractor, D., McDuff, D., Haines, J. K., ... & Li, H. (2022). Behavioral use licensing for responsible AI. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 778-788). https://doi.org/10.1145/3531146.3533143.

Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., ... & Gardner, M. (2021). Documenting large webtext corpora: A case study on the colossal clean crawled corpus. arXiv preprint arXiv:2104.08758. https://doi.org/10.48550/arXiv.2104.08758.

Eldan, R., & Li, Y. (2023). TinyStories: How Small Can Language Models Be and Still Speak Coherent English?. arXiv preprint arXiv:2305.07759. https://doi.org/10.48550/arXiv.2305.07759.

Liang, P., Hashimoto, T., Ré, C., Bommasani, R., Xie, S.M. (2024). Data. CS324 - Large Language Models. https://stanford-cs324.github.io/winter2022/lectures/data/ (consulté le 06/03/2024).

Dignum, V. (2019). Responsible artificial intelligence: how to develop and use AI in a responsible way (Vol. 2156). Cham: Springer. ISBN : 978-3-030-30373-0.

Ferrara, E. (2023). Should chatgpt be biased? Challenges and risks of bias in large langage models. arXiv preprint arXiv:2304.03738. https://doi.org/10.5210/fm.v28i11.13346.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30, 681-694. https://doi.org/10.1007/s11023-020-09548-1.

IEA (2024). Data Centres and Data Transmission Networks. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks (consulté le 12/02/2024).

Jullien, N., Viseur, R., Zimmermann, J.-B. (2022). Gouvernance d'un projet libre : contrôler un flux d'innovation. Enjeux numériques, juin 2022, n°18. https://www.annales.org/enjeux-numeriques/2022/en-2022-06/2022-06-13.pdf.

Jullien, N., & Viseur, R. (2021). Les stratégies open-sources selon le paradigme des modèles économiques. Systèmes d'Information et Management, 26(3), 67-103. https://doi.org/10.3917/sim.213.0067.

Pénin, J. (2011). Open source innovation: Towards a generalization of the open source model beyond software. Revue d'économie industrielle, (136), 65-88. https://doi.org/10.4000/rei.5184.

Riehle, D. (2010). The economic case for open source foundations. Computer, 43(01), 86-90. https://doi.ieeecomputersociety.org/10.1109/MC.2010.24.

Shaikh, M., & Cornford, T. (2011). Total cost of ownership of open source software: a report for the UK Cabinet Office supported by OpenForum Europe. https://eprints.lse.ac.uk/39826/1/Total_cost_of_ownership_of_open_source_software_(LSERO).pdf.

Sundberg, N. (2024). Tackling AI's Climate Change Problem. MIT Sloan Management Review, 65(2), 38-41. https://sloanreview.mit.edu/article/tackling-ais-climate-change-problem/.

West, J. (2003). How open is open enough?: Melding proprietary and open source platform strategies. Research policy, 32(7), 1259-1285. https://doi.org/10.1016/S0048-7333(03)00052-0.

# Automatic Categorization of ESWD Weather Reports in French

**Mija Pilkaite** [1], **Davide Buscaldi** [2]

1. *LIX, Ecole Polytechnique*
   *91120 Palaiseau, France*
   *mija.pilkaite@polytechnique.edu*

2. *LIPN, CNRS UMR 7030*
   *Université Sorbonne Paris Nord*
   *93430 Villetaneuse, France*
   *buscaldi@lipn.fr*

RÉSUMÉ. *Dans le contexte des préoccupations liées au changement climatique, les phénomènes météorologiques violents représentent une problématique majeure en raison de leur incidence sur la société humaine. Les chercheurs ont collecté des données sur ces événements afin de mieux comprendre leur corrélation avec le changement climatique et d'améliorer notre capacité à les prédire et à nous y préparer. Le Laboratoire européen des tempêtes violentes (ESSL) a mis en place la base de données européenne sur les phénomènes météorologiques violents (ESWD), permettant au public de signaler des évenements de ce type. L'utilisation du traitement automatique de diverses sources médiatiques, telles que les actualités et les médias sociaux, a suscité un intérêt croissant afin d'identifier et de recenser les phénomènes météorologiques violents de manière plus précise et objective. Ce travail se concentre donc sur l'utilisation de différentes techniques pour extraire les informations pertinentes et rendre l'ESWD moins dépendant de l'intervention humaine.*

ABSTRACT. *Amidst the concerns for climate change, severe weather events represent an important issue because of their impact on human society. Researchers have been collecting data regarding these kinds of events to try to understand their relationship to climate change and improve our ability to predict and prepare for them. The European Severe Storms Laboratory has created the European Severe Weather Database (ESWD), which allows the public to report and share information about severe weather events. To address this issue, there has been growing interest in using automatic processing of various media sources, such as news and social media, to identify and survey severe weather events more accurately and objectively. Thus, this work focuses on using different techniques to extract the relevant information and make the ESWD less human-dependent.*

MOTS-CLÉS : *European Severe Weather Database, Classification, TAL*

KEYWORDS: *European Severe Weather Database, Classification, Natural Language Processing*

## 1. Introduction

As the advancements in Machine Learning have been fast-paced, we are looking for more adaptations and uses for it in the science world. Climate change has been also a burning topic for several years, thus this project provides an opportunity to connect these two fields to better explore the severe weather events in Europe. The analysis and prediction of severe weather events are of great importance, as extreme weather conditions can lead to significant economic and human losses (C. A. Doswell, Kay, 2005) and accurate categorization and extraction of severe weather events can help understand the frequency, intensity, and distribution of such disasters, as well as for planning and designing effective mitigation strategies (Jessica Mercer, Taranis, 2009). The European Severe Weather Database (ESWD) serves as a valuable resource for weather data, containing reports of severe weather events in Europe within most of the countries and a wide range of categories (Groenemeijer *et al*., 2009). The ESSL [1] (European Severe Storms Laboratory) started as an informal network of European scientists to advance research on severe convective storms and extreme weather events on a European level; today it is managing an actively growing database that could significantly benefit from automation, particularly in incorporating events sourced from news and social media. This idea is at the core of this preliminary study - assessing the viability of leveraging Natural Language Processing techniques to discern whether a news article references a severe weather event worthy of inclusion in ESWD, and in the affirmative case what would be the category of the event.

## 2. Data Description and Preprocessing

We obtained from ESSL a set of 22,317 entries from the ESWD, covering severe weather events that occurred in France from January 1, 2016, to December 31, 2022. For each event, 22 fields are present: among these a unique id, timestamp, latitude and longitude, location, region, meteorological data, number of victims and injured, the type of the event, an event description, and a reference (usually the title and the link to an on-line article). Since we wanted to predict the event type from text, we used the 'EVENT TYPE' column as target and the 'REFERENCE' column for the input text (the event description is often empty as it's an optional short description added to the event). As it is shown in Figure 1, there are 8 different severe weather events and the dataset is unbalanced.

However, of the total 22,317 entries, only 12,867 contain a non-empty reference, reducing the set of available data. We also had to preprocess the data to remove empty or non-informative references, such as the following ones:

*"Joel Feneule (on Facebook), 17 Nov 2022."*
*"Report via Kachelmannwetter.com, 4 Nov 2022."*

1. https://www.essl.org/cms/

Finally, we used Spacy [2] language detection to keep only the texts in French. After this step, we obtained a set of 3,246 reports. More pre-processing was required to clean reports such the following one:

*"Eyewitness report via Alpes 1 (on Facebook), 14 SEP 2022. ""Hautes-Alpes : un violent orage de grêle s'est abattu sur Gap"*

in order to keep only the second part of the reference. This was done with hand-made rules (for instance, removing all text matching patterns such as "eyewitness report" or texts that are below a length threshold of 5 words) and the use of Spacy's language detection features.

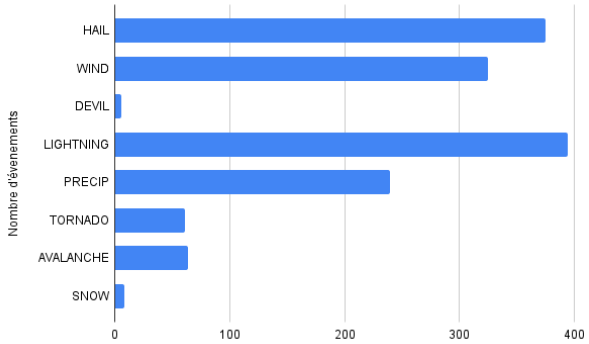



FIGURE 1. *Distribution of the 8 event types in the final dataset.*

At the end of this pre-processing step, we kept $1,472$ events distributed as seen in Figure 1. As it can be seen, the wind category is not so dominant as in the full one. The reason is that many of the wind reports come from meteorological station reports, therefore there is no meaningful text associated.

## 3. Text Classification Models

We classified the text focusing on various types of text representations and classification algorithms.

**Text Representation** We represented the text of the reports in three different ways: as Bag-of-Words with the frequency of the words, tf.idf (term frequency-inverse document frequency) vectors, and dense vectors obtained using Sentence-BERT. For both BoW and tf.idf we remove stopwords and keep the words that occur at least 2 times. In this way, reports are represented by sparse vectors of size $2,918$. Sentence-BERT (SBERT) is a sentence encoder based on Siamese architecture, which modifies the

2. https://spacy.io/

BERT (Bidirectional Encoder Representations from Transformers) architecture, showing state-of-the-art performance in various natural language processing tasks (Devlin *et al.*, 2018). SBERT can be used to encode each sentence into a dense vector of size 768.

**Classification Methods** We considered three widely used algorithms: logistic regression (LR), random forests (RF) and a fully connected neural network (FCNN). For the random forest classifier we employed hyperparameter tuning using GridSearch with 3-fold cross validation. The FCNN was set up with two hidden layers of 512 units each both with Dropout layer (dropout rate 0.2) and a Softmax output with categorical cross-entropy loss. We also set the batch size at 64 and 20 training epochs. We considered fine-tuning a BERT model, but we faced problems due to the imbalance of the dataset (the model was predicting always the same label).

## 4. Results and Analysis

We split the data into a random partition, using 80% of data for training and 20% for testing. The results across all representations and methods are shown in Table 1.

TABLEAU 1. *Accuracy obtained using the various text representation methods and classification methods.*

| Text Representation | LR | RF | FCNN |
|---|---|---|---|
| BoW | 82.9% | 68.9% | 82.6% |
| tf.idf | 83.9% | 68.7% | 81.6% |
| SBERT | 82.1% | 67.8% | 75.1% |

From the results, it is evident that Sentence-BERT is not adequate to represent the data in this kind of task. Bag-of-Words is a representation that on average yields the best results across all methods. Tf.Idf is only better with the LR classifier. This result can be explained by the fact that the rare words (that are boosted by idf) are not important clues for understanding the category of the event. Random Forests performed poorly on the dataset. From the confusion matrix in Figure 2, it can be seen that this model is never able to predict the PRECIP class, being split between LIGHTNING and WIND. As it is evident from Figure 3, the discriminative power of single words tends to be weak and there is a bias towards the most represented classes.

We examined the best results to identify the situations that still posed problems for classification. First of all, we calculated the confusion matrix in Figure 4. From this matrix, it can be seen that PRECIP is the category less easily identified as it is often confused with WIND.

Let us look at some misclassified examples:

Actual: LIGHTNING Predicted: WIND Reference: ['Orages', 'départ', 'feu', 'maison', 'détruite', 'arbre', 'couché', 'tarn', 'Haute-Garonne', 'tarn-et-garonne', 'FRANCE', 'TV', 'INFO', '30', 'aug', '2022']

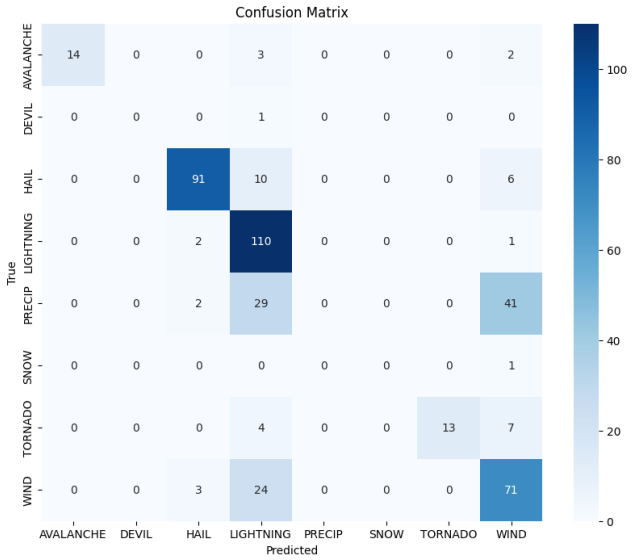



FIGURE 2. *Confusion matrix for tf.idf and Random Forests*

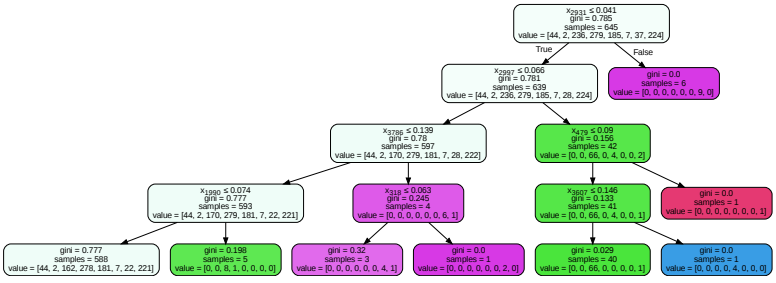



FIGURE 3. *A single Decision Tree from the RF model with tf.idf weights*

Actual: HAIL Predicted: WIND Reference: [’violent’, ’orage’, ’évacuation’, ’Saint-Etienne’, ’femme’, ’prisonnier’, ’voiture’, ’Villars’, ’Progrès’, ’01’, ’Jul’, ’2019’]

Actual: LIGHTNING Predicted: HAIL Reference: [’intempérie’, ’orage’, ’faire’, ’gros’, ’dégât’, ’ouest-aveyron’, ’vendredi’, ’soir’, ’ladepeche.fr’, ’29’, ’june’, ’2020’]

Most of the misclassifications stem from words that can be used for multiple events. For example, the word ’orage’ for the model is typically associated with wind, but it can be also associated to lightning and hail as we can see in the first two examples.

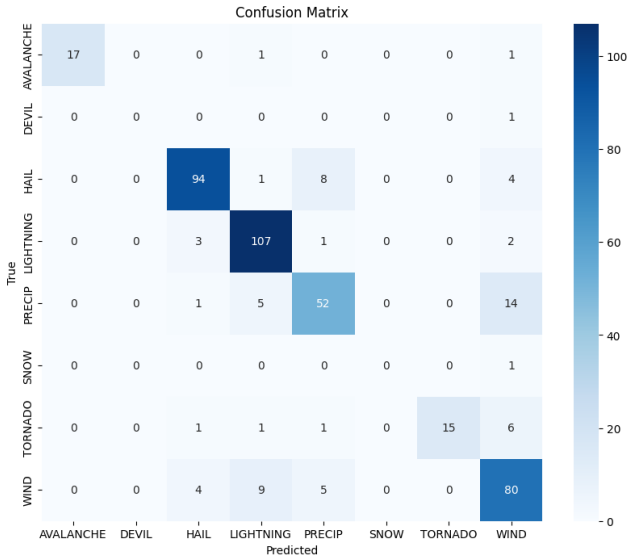



FIGURE 4. *Confusion matrix for tf.idf and Logistic Regression*

We then extracted from the LR model the most important features (words) for each target class. In Table 2, we sort the top 10 features by their weight magnitude in descending order.

TABLEAU 2. *Most important features for each event type.*

| | AVALANCHE | DEVIL | HAIL | LIGHTNING | PRECIP | SNOW | TORNADO | WIND |
|---|---|---|---|---|---|---|---|---|
| 1 | avalanche | tourbillon | grêle | foudre | inondation | neige | tornade | ws |
| 2 | skieur | jul | grêlon | incendie | orages | 15 | 2019 | meteofrance |
| 3 | avalanch | apr | limousin | foudroyer | inonder | jan | oct | arbre |
| 4 | jan | jardinier | jun | kachelmannwetter | boue | 000 | facebook | mini |
| 5 | savoie | ferté | agriculteur | jun | pluie | souffert | nov | feb |
| 6 | 2021 | bernard | eyewitness | sep | eau | le | dec | tempête |
| 7 | dec | surprendre | report | maison | provence | lorrain | direct | vent |
| 8 | alpes | ouest | centre | nord | oise | alsace | keraunos | coup |
| 9 | mort | tornad | bilan | feu | orage | chute | youtube | report |
| 10 | feb | argelès | 21 | com | progrès | priver | azur | march |

As we have seen through the process, there are various keywords that help our models identify which severe weather event the reference is talking about. However, this also leads us to another pressing question - how do we make that our models do not confuse references about not-weather events or about weather events that are not severe if they contain these keywords? This is explored in the final part of this work.

## 5. Binary classification of Severe Weather events

Besides classifying the event types, it is crucial to have a classifier that could distinguish between severe weather events and not or even other articles that contain

similar keywords. For this reason, we scraped from the web 500 headlines containing some of the most important keywords seen in Table 2, and built a classifier with these headlines (negative class) and 500 randomly picked events from the ESWD database, labeled as positive samples.

With this setup and 10-fold cross-validation, we obtain for tf.idf and LR $98.9\%$ accuracy, indicating that it is possible to effectively diversify the ESWD-worthy reports from general news related to weather.

Some examples of misclassified instances:

– Haute-Savoie glissement terrain bloque route Thyez dauphiner Libéré 15 July 2021 for suscriber only (True label: 1, Predicted label: 0)

– grêle sud-est orage l'Ouest (True label: 0, Predicted label: 1)

– météo France surprenante (True label: 0, Predicted label: 1)

– Mickael B. observatoire ciel Orageux Tornade Médoc 2018 2019 (True label: 1, Predicted label: 0)

As it can be seen, for the second example the model is probably correct as we didn't filter out those headlines referring to severe weather. In the other cases, the context is not big enough, which represents probably the main difficulty of this task.

## 6. Conclusions

In this work, we explored the classification of severe weather events using textual data from the European Severe Weather Database, using various classification models. Overall, most of the models tested achieved good performance with $\sim 80\%$ accuracy, indicating the textual data contains meaningful signals to distinguish between different types of severe weather events; although the imbalance of classes poses an obstacle to further improvement of these results. We have also seen that the data collected by volunteers is especially noisy and is not held to any standard, thus making any sort of processing quite a difficult task. We also developed a binary classifier to filter out false data that contains key weather-related terms but does not actually describe a severe weather event. This classifier achieved over 98% accuracy, indicating that it is possible to create a severe weather monitor for news and social media to enrich the ESWD with automatically collected information.

There are several promising avenues for future work based on this project. With a larger dataset, deep learning models may achieve even higher accuracy in classifying severe weather events and sub-categories. Contextualized word embedding models like BERT (Devlin *et al.*, 2018) and RoBERTa (Liu *et al.*, 2019) should also be explored as they have achieved state-of-the-art results on various text classification tasks.

## Bibliographie

C. A. Doswell H. E. B., Kay M. P. (2005). Climatological estimates of daily local nontornadic severe thunderstorm probability for the united states. *Weather and Forecasting*, vol. 20, n° 4, p. 577-595.

Devlin J., Chang M.-W., Lee K., Toutanova K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Groenemeijer P., Kuehne M., Liang Z., Dotzek N. (2009). New capabilities of the european severe weather database. *In 5th European conference on severe storms, Landshut, Germany*, p. 311-312.

Jessica Mercer I. K., Taranis L. (2009). Framework for integrating indigenous and scientific knowledge for disaster risk reduction. *Climate Change Modeling, Mitigation, and Adaptation*, vol. 34, n° 1, p. 214-239.

Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D. *et al*. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.