

Projet de thèse GouD "Gouvernance des données dans le contexte ESRI"

Appel à candidatures

Encadrement scientifique du projet de thèse.

Guy Melançon, PR , section 27 (Informatique) guy.melancon@u-bordeaux.fr

Pinède, Nathalie, MCF HDR, section 71 (Sciences de l'information et de la communication) nathalie.pinede@u-bordeaux-montaigne.fr

Descriptif général

Les laboratoires LaBRI (UMR 5800) et IMS (UMR 5218) de l'Université de Bordeaux recherchent un candidat pour travailler sur le thème de la gouvernance des données en contexte universitaire. Ce travail s'inscrit dans le cadre du projet d'établissement ACT visant à développer, tester, valider et diffuser de nouvelles approches aux grands problèmes environnementaux, sociaux et de transition économique faisant des campus de l'université un vaste laboratoire vivant. La thèse constitue un volet du projet GouD (Gouvernance des Données) lui-même partie prenante de ACT (*Augmented university for Campus and world Transition*), dont la gouvernance des données est une priorité, dans un contexte de science ouverte et de pilotage par les données des institutions publiques.

La thèse se penchera sur des questions dont s'empare aujourd'hui le domaine de la recherche en système d'information, mais qui restent pour beaucoup sans réponse dans le contexte particulier de l'enseignement supérieur et de la recherche française. Le projet GouD compte définir les mécanismes de cette gouvernance, allant jusqu'à identifier les acteurs et leurs responsabilités. Cette gouvernance devra à l'évidence s'articuler avec d'autres instances ou gouvernances (RGPD, PSSI, données de santé, obligation légale d'archivage...). Partant de travaux récents de la communauté de recherche dans le domaine des systèmes d'information, GouD se propose de travailler depuis le terrain en ancrant sa réflexion au niveau des *living labs* labellisés ACT, afin de co-construire une gouvernance adaptée à chacun de ces projets, faisant de ces propositions le socle d'une gouvernance au niveau de l'établissement.

Le/la candidat.e recherché.e, de profil informatique et/ou sciences de l'information, doit montrer un goût pour les questions touchant aux données et aux systèmes d'information, tant d'un point de vue technique qu'organisationnel. Le déroulement du projet amènera le candidat à échanger avec des acteurs aux profils différents, chercheurs impliqués dans des projets soulevant des questions de gouvernance des données ou endossant une responsabilité au niveau institutionnel, par exemple.

Les candidat.e.s intéressé.e.s sont invités à prendre contact avec les 2 co-encadrants **avant le 15 octobre 2022** (CV et lettre de motivation demandés).

Contexte.

Les données sont souvent mises en avant comme potentiel vecteur de développement des organisations. Elles sont présentées comme outil compagnon du pilotage d'une organisation et de la construction de sa stratégie (Waller 2020). La littérature vantant les mérites du « pilotage par les données » est volumineuse, mais emprunte le plus souvent un style qui relève de la vulgarisation scientifique voire de la « propagande » destinée aux acteurs empressée de ne pas rater le rendez-vous d'une transformation qu'il peine à comprendre (Ernst & Young – Mourroux & Deballon 2021), McKinsey (Assur & Rowshankish 2022)...

Dans ce contexte, les aspects de gouvernance des données sont occultés par les questions plus opérationnelles de fouille et d'analyse des données afin d'éclairer la décision; ce sont en effet les usages des données qui intéressent la plupart des acteurs. Bien que la littérature sensibilise au besoin et au bénéfice de doter les organisations d'une gouvernance des données, la définition même de ce qu'est cette gouvernance, sa fonction et sa mise en place sont rarement décrites.

Cela dit, des initiatives sont prises par des organisations afin d'avancer sur le sujet; des définitions, empreintes de pragmatisme en émergent. Par ailleurs, le contexte réglementaire de l'ESRI (Enseignement Supérieur, Recherche et Innovation) apporte des éclairages, mais encore une fois plutôt sur les usages et sur les aspects qui concernent le système d'information, et dans un cadre beaucoup plus large que le seul contexte universitaire (« Cadre Commun d'Urbanisation du SI de l'Etat », v1.0, 2012) (Rapport au Premier Ministre 2016-2017).

La déclinaison du cadre réglementaire au niveau des universités n'est pas immédiate, compte tenu de l'autonomie dont elles disposent en matière stratégique, et des spécificités de leurs missions. La définition de la gouvernance des données, de son organisation et de son fonctionnement dans un établissement universitaire reste à notre sens à faire.

Selon (Abraham et al., 2019), la gouvernance doit viser à faciliter et fluidifier le parcours des enseignants-chercheurs et des services au travers des différents cadres dont relève les usages de la donnée. L'organisation mise en place par la gouvernance doit permettre à l'établissement de mener une réelle politique en matière de données, sur tous les domaines où l'amènent ses missions.

Les mots clés associés à la gouvernance expliquent à eux seuls sa raison d'être:

- Les données doivent être fiables dès lors qu'il s'agit de s'y appuyer en matière stratégique et décisionnelle.
- Les responsabilités en matière de fiabilité, de sécurité ou encore de disponibilités des données doivent être clairement définies.
- C'est à la gouvernance qu'il revient de définir, sur certains périmètres, une politique et sa traduction en termes de règles et procédures, de standardisation. La définition et la conduite d'une feuille de route d'une stratégie « *data-driven* » en dépend.

Dans le contexte universitaire, les données arrivent de plusieurs scènes. C'est celle de la recherche, aujourd'hui invitée à suivre la voie de l'ouverture (*open science*). C'est celle de la formation, que certains espèrent utiliser pour seconder les équipes pédagogiques à des fins d'orientations ou d'encadrement des étudiants (*learning analytics*). C'est celle de l'organisation et de ses nombreux processus internes, dont on espère construire un jumeau numérique (*data-driven management/strategy*).

Cette description suggère une partition des données en classes disjointes, mais qu'en est-il ? Peut-on en effet aborder les données de manière segmentée, en définissant une gouvernance différenciée pour chacun de ces « types » de données ? Comment conjuguer une telle approche avec les cadres réglementaires comme le RGPD sur la protection des données à caractère personnel, la réglementation propre aux données de santé, ou encore la PSSI (politique de sécurité des systèmes d'information) qui s'appliquent de manière transversale à l'établissement ?

La gouvernance des données suit des ambitions d'un « pilotage par la donnée » de l'établissement, et des objectifs affichés de la politique d'ouverture des données de la recherche. La démarche exige de définir les principes d'une gouvernance guidant toutes initiatives produisant, consommant ou manipulant des données, quelle que soit leur nature, et une mise en cohérence d'initiatives se déployant sur toutes les missions de l'établissement et se déroulant à différentes échelles.

Incidemment, le développement de *living labs* verra naître des besoins de gestion de données propres à ces projets et constitue en quelque sorte un objet de recherche privilégié pour avancer sur le sujet de la gouvernance. Le lancement d'un projet de recherche action développant les fondements d'une gouvernance des données de l'établissement vient donc répondre au besoin d'une meilleure gestion du cycle de vie des données de ces initiatives, avant que de servir de socle pour définir une gouvernance au niveau de l'établissement.

Objectifs

L'objectif de cette recherche action est d'énoncer clairement les fondements de la gouvernance des données dans le contexte ESRI, et de produire une méthodologie robuste permettant à un établissement universitaire de se doter d'une telle gouvernance. Un corollaire immédiat est la définition au niveau de notre établissement d'une gouvernance des données adaptée à son contexte. La généralisation à tout établissement de l'ESRI, ou à tout le moins aux établissements de profils similaires (membres de UDICE, par exemple) pourra suivre.

Les approches de gouvernance des données sont encore peu documentées, et lorsqu'elles le sont, elles concernent le plus souvent le monde de l'entreprise. La complexité du contexte ESRI en fait une question complexe, certainement due à la multiplicité et l'enchevêtrement des missions de l'établissement, et la multiplicité des acteurs dans le contexte et l'écosystème universitaire.

Ainsi, en associant les perspectives des domaines des systèmes d'information et de la gouvernance de l'information, l'ambition est de proposer un fondement à la gouvernance des données pour un établissement ESRI.

Cela exige de dessiner les mécanismes propres à cette gouvernance, allant jusqu'à identifier les acteurs et leurs responsabilités. On peut s'attendre à une articulation de la gouvernance sur plusieurs niveaux, mettant en cohérence les différents projets consommant ou produisant des données avec les principes de

la gouvernance au niveau établissement. Par ailleurs, cette gouvernance devra à l'évidence s'articuler avec d'autres instances ou gouvernances (RGPD, PSSI, données de santé, obligation légale d'archivage, ...).

Méthodologie

Nous envisageons d'amorcer ce travail de recherche sur la base de la synthèse de (Abraham et al. 2019). Ce travail propose un cadre méthodologique qu'il convient de confronter au contexte de l'établissement.

La démarche initiale consisterait à « confronter » chacun des projets à ce cadre méthodologique afin à la fois d'aider les projets à structurer leur démarche de gouvernance, mais aussi de poser un regard critique sur la méthode elle-même. Resterait ensuite à énoncer les principes, et à formuler des propositions, pour une gouvernance des données au niveau de l'établissement, émergeant des différents cas d'usage étudiés.

Le cadre méthodologique de (Abraham et al. 2019) structure la gouvernance des données autour de mécanismes de différents types:

- relationnels, visant principalement l'adoption de la démarche et l'accompagnement au changement ;
- structurels, désignant les responsabilités (et les responsables) sur les différentes phases du cycle de vie de la donnée, des « producteurs » vers les « consommateurs » ;
- procéduraux, chargé de la définition et l'implémentation d'une politique de la donnée, aligné avec les politiques connexes (données personnels, PSSI, ...) et avec la stratégie globale de l'organisation.

Toujours selon (Abraham et al. 2019), ces mécanismes agiront de manière différenciée selon les cas et soulèveront des questions à résoudre se déclinant sur trois facettes:

- organisationnelle, où entrent typiquement en jeu les relations partenariales, dès lors qu'un projet sort des frontières de l'établissement ;
- physique (ou support des données), où sont prises en compte la nature des données (SI, référentiel, transactionnelle, données d'analyse, traces d'activités, ...) et leur origine (capteurs, données d'enquête, document d'archives, données administratives, ...) ;
- décisionnelle, incluant les questions de qualité des données (appelant une démarche managériale spécifique), de sécurité (confidentialité, l'authenticité, l'accessibilité, la disponibilité, l'accès contrôlé...), et plus généralement du cycle de vie de la donnée, où sont typiquement abordées les questions relatives aux infrastructures IT.

Résultats attendus, retombées et valorisation

L'impact du projet GouD se fera sentir déjà au niveau des *living labs* et projets qu'il accompagnera. Un résultat central du travail doit être la définition et une première mise en place de la gouvernance au niveau établissement (sur un périmètre englobant ACT), en lien avec l'« échelon projet ».

Des retombées sont attendues, comme c'est le cas pour tout projet de recherche, en termes de production scientifique (articles en revue/conférence). La dissémination des résultats du projet, notamment de la méthodologie qui sera produite, pourra se faire au travers de séminaires dédiés auprès de tous les acteurs de l'établissement.

Les résultats pourront intéresser, au-delà de la communauté scientifique, la communauté ESRI et les acteurs institutionnels proches du sujet tels l'AMUE (Agence de Mutualisation des Universités et Établissements), et plus largement les acteurs soucieux de l'organisation du service public de la donnée.

Références.

- Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49 (August), 424–438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Assur, N., Rowshankish, K. (2022). *The data-driven enterprise of 2025*. McKinsey (January 2022) URL: <https://www.mckinsey.com/business-functions/quantumblack/our-insights/the-data-driven-enterprise-of-2025>
- Mourroux, J., Deballon, D. (2021). *La donnée et sa gouvernance, un enjeu majeur pour le pilotage et la prise de décision*. Ernst & Young, URL: https://www.ey.com/fr_fr/government-public-sector/la-donnee-et-sa-gouvernance (consulté le 31/5/2022).
- Waller, D. (2020). *10 steps to creating a data-driven culture*. Harvard Business Review 6. <https://hbr.org/2020/02/10-steps-to-creating-a-data-driven-culture>
- Rapport au Premier Ministre sur la donnée dans les administrations, 2016-2017 : *La donnée comme infrastructure essentielle*.
- *Cadre Commun d'Urbanisation du SI de l'Etat*, v1.0, 2012.