



Les séminaires de connaissance réciproque : et si nous faisons connaissance ?

Les séminaires de connaissance réciproque sont l'occasion de rencontrer nos futurs collègues pour échanger sur différents sujets concernant notre nouvelle université, que ce soit sur des sujets scientifiques, de formation, de support et de soutien.

Séminaire des 01 et 02/02/2018, ENSG

BILAN ET RESTITUTION DE L'ATELIER

Enjeux du numérique – Déluge de données

Le numérique est un enjeu majeur pour la construction de l'université cible. La transformation numérique introduit des changements majeurs de paradigmes qui bouleversent en profondeur à la fois nos façons de travailler mais aussi nos objets d'étude. Dans ce contexte en évolution permanente, il est indispensable d'être capable de détecter et de saisir les opportunités : c'est une chance pour renouveler nos approches et nos fonctionnements, pour moderniser nos outils, pour mieux travailler ensemble, pour mieux nous faire connaître, pour mieux exploiter nos données afin de mieux cibler nos actions... Sans cela, on a peur d'être ringardisés, d'être disruptés ou pire ... d'être uberisés.

L'atelier « Enjeux du numérique – Déluge de données » est conçu en dyptique : la première partie qui s'est tenue à l'ENSG s'est focalisée sur le thème des **données**, tandis que le second volet, qui aura lieu en avril à l'ESIEE sera centré sur l'impact du numérique sur nos organisations et nos métiers.

Ce premier atelier a réuni 20 personnes en provenance de l'ENSGE (5), de l'ESIEE (1), de l'Ifsttar (11) et de l'UPEM (3). Les participants avaient des profils variés (chercheurs, ingénieurs, valorisation, IST, enseignants, administratifs, communication) et des expériences différentes propres à leur métier sur le thème de l'atelier.

Les participants de l'atelier ont été invités dans un premier temps à partager leur expérience sur les pratiques de traitement de données au sein des différents établissements. Dans un second temps, à partir d'une analyse du constat réalisé, des pistes d'améliorations ont été collectivement élaborées et débattues pour aboutir à une série de quelques actions phares qui pourraient être développées dans le cadre du projet Future.

Les participants :

Jean-Michel AUBERLET (Ifsttar), Mourad BEN HADJ (UPEM), Arnaud BONNARD (Ifsttar), Mathieu BREDIF (ENSG), Elba BURITY (Ifsttar), Catherine DOMINGUES (ENSG), Éric DUMONT (Ifsttar), Monssef DRISSI-HABTI (Ifsttar), Michèle GUILBOT (Ifsttar), Jean-François HANGOUËT (ENSG), Laurent JAIL (UPEM), Aghiad KHADOUR (Ifsttar), Patrick LACOUR (Ifsttar), Emilie MAUCOURT-BACCHI (Ifsttar), Jean-Pierre PAPELARD (ENSG), Antoine PINTE (ENSG), Didier RICHARD (ENSG), Pascal ROMON (UPEM), Céline ROUSSELOT (Ifsttar), Rochdi TRIGUI (Ifsttar), Emilie VIDAL (Ifsttar)

« C'est quoi les données » ?

Lors du premier temps de l'atelier, chaque participant était invité, par le biais d'un tour de table, à présenter succinctement les données sur lesquels il était amené à travailler dans le cadre de ses activités.

Suite à ces échanges, un travail collectif de réflexion¹ a permis de formaliser les éléments de réponse de la manière suivantes :

Sources de données (tous métiers confondus : recherche, administration ...)	<ul style="list-style-type: none">• Données subjectives, données d'enquête - l'idée qu'on se fait du monde• Données objectives, données capteurs - le modèle numérique du monde• Données fabriquées, données issues de codes sources - une création numérique (idéalement) conforme au monde• Les codes sources sont aussi des données• Données externes, de référence (ou pas!)• Données implicites et émergentes (par exemple, des données de gestion périphériques aux activités)
Traitements effectués	<ul style="list-style-type: none">• Nettoyage ou « apurement » des données : objectif qualité• Extraction de l'information (avec des outils... ou avec les yeux!)• Post-traitements (avec possible retour à une étape d'extraction de l'information)
Destinations, usages	<ul style="list-style-type: none">• Usagers cibles• Actions - buts• Canal de diffusion <p><i>Nb : les contributions étaient tellement variées pour cette dimension qu'il a fallu beaucoup agréger, d'où les catégories très générales.</i></p>

En synthèse, ce premier état des lieux nous démontre une grande variété et une grande hétérogénéité aussi bien sur les sources données, sur les traitements effectués et encore plus sur les usages.

Au-delà de ces résultats, au fil des discussions, certains mots-clés ont émergé à de multiples reprises. On peut notamment mentionner les mentions récurrentes des participants sur les questions de :

- sécurité, confidentialité
- stockage, archivage, conservation
- puissance de traitement

¹ Le nuage de post-it produit est téléchargeable à l'adresse : <https://mycore.cnrs.fr/index.php/s/X0qvYkhxSHvPTEC/download?path=%2F&files=nuage%20de%20post-its.pdf>

- classification, documentation
- diffusion, partage, accès (cf. FAIR²)
- visualisation
- masse, filtrage
- juridique, protection des personnes, anonymisation, responsabilité
- transparence des algorithmes
- homogénéisation des pratiques et politique de la donnée (par exemple, usage de DMP³)
- code source : une donnée (presque) comme les autres !

Le traitement et la gestion des données : satisfaction, soucis et blocage

Lors du second temps de l'atelier, chaque participant était invité, toujours par le biais d'un tour de table, à présenter la manière dont ses données étaient traitées.

Suite à ces échanges, un travail de réflexion a cherché à mettre en avant 3 aspects liés au traitement des données :

- Les satisfactions : « *Ce qui marche* »
- Les besoins : « *Ce qui manque* »,
- Les obstacles : « *ce qui bloque* »

	Satisfactions « <i>Ce qui marche</i> »	Besoins « <i>Ce qui manque</i> »	Obstacles « <i>ce qui bloque</i> »
Outils	Des outils fonctionnels <ul style="list-style-type: none"> • Partage • Numérisation / capacité à la collecte de données • Moyens fournis et volume disponible de stockage 	Des outils à améliorer <ul style="list-style-type: none"> • Infrastructure de stockage et d'accès aux données • Outils de partage pour gros volume, sécurisé, souverain • Archivage, sécurisation à long terme 	Des outils manquants <ul style="list-style-type: none"> • Intégration des outils au SI • Peu de doc décrivant les SI • Archivage intégré au SI • Espace de stockage • Alternatives aux outils Dropbox, Google et autres
Méthodes	Des méthodes appropriées <ul style="list-style-type: none"> • Volonté des chercheurs à mieux gérer les données 	Des méthodes à généraliser <ul style="list-style-type: none"> • Guides de bonnes pratiques (y compris le code) • Normalisation, 	Des méthodes à promouvoir <ul style="list-style-type: none"> • Qualifier la qualité des données • Formats ouverts au lieu de

² FAIR correspond à « Findable, Accessible, Interoperable, Reusable », principes sur lesquels repose la stratégie d'ouverture des données de recherche dont la commission européenne fait la promotion. À rapprocher aussi de l'initiative Responsible Data Science (Cf. <http://www.responsibledata.org/>) avec le « Fairness, Accuracy, Confidentiality, Transparency » sur les mentions récurrentes.

³ DMP signifie Data Management Plan, ou plan de gestion de données. La rédaction, tout au long de la vie d'un projet, d'un plan de gestion des données formalisant l'ensemble des activités liées aux données est désormais considéré comme une bonne pratique de recherche.

	<ul style="list-style-type: none"> Formalisation commune 	<ul style="list-style-type: none"> standardisation (formats communs et standards) Transparence (sur les données existantes et sur les traitements réalisés) Généraliser l'usage de plan de gestion de données (formaliser) Exploitation plus systématique des données (des étudiants...) Temporalité et gestion temporelle des versions des données Partage de données, de code 	<ul style="list-style-type: none"> propriétaires Informer pour Dropbox, Google et autres risques Efforts pour respecter la législation contraignante pour protéger les droits des personnes Pas d'anticipation sur la gestion des données, pas de plan de gestion
Dimension institutionnelle	implication institutionnelle <ul style="list-style-type: none"> Politique Communication interne Portage du sujet « données » via un travail pluridisciplinaire 	Une prise de conscience institutionnelle à développer <ul style="list-style-type: none"> Du temps !!! Des sous!!! Des gens!!! Politique de gestion des données Compétence « data » d'un comité d'éthique Manque de connaissances, de compétences Prendre en compte la variété de la réglementation à l'échelle internationale Accompagner la diffusion 	Des verrous institutionnels <ul style="list-style-type: none"> Connecteurs entre silo (inter et intra établissement(s)) Moyens Formation tout au long de la carrière : Augmentation des compétences techno Besoin de conseil, d'accompagnement Formation pour les chercheurs sur leurs droits et obligations Résistance au changement et aux procédures

Pour cet état des lieux, il est important de rappeler que les outils, les méthodes et l'investissement institutionnel sont différents d'un établissement à l'autre. Cela explique donc pourquoi certains éléments sont classés à la fois dans la colonne des satisfactions (lorsqu'un établissement aborde bien le sujet) et dans les colonnes des blocages (lorsqu'un établissement ne propose pas de solution sur le sujet).

Ces contenus sont intéressants et montrent les trois niveaux de la problématique : des questions à résoudre au niveau opérationnel (des outils), au niveau tactique (des méthodes) et au niveau stratégique (des institutions).

Actions proposées pour l'université cible

La réflexion collective du groupe a abouti, lors du troisième et dernier temps de l'atelier, à la liste suivante d'actions phares qui pourraient être développées dans le cadre de l'université cible, afin d'élaborer une **stratégie numérique collective** :

- **Diagnostiquer** / formaliser le traitement des données
 - État des lieux / cartographie des données / Classification
 - État de l'art et parangonnage
- **Construire une stratégie numérique collective**
 - Définir une politique numérique commune (alignée avec les autres stratégies : valorisation, ...)
 - Identifier des chantiers numériques (par exemple, learning analytics, suivi des alumni ...)
 - Mettre en œuvre des méthodes innovantes (par exemple, machine learning ...)
- Renforcer la **politique de responsabilité** sur les données
 - Comité d'éthique UCible compétent sur les données
 - Connaissances juridiques (données personnelles, PI ...)
- **Outiller**
 - Offre interne, ouverte vers l'extérieur
 - Pour le partage, pour le chiffrage, pour la pseudonymisation, pour l'archivage ...
- **Former, informer, accompagner** : action en continu
- Mesurer et auditer en continu la mise en œuvre de la stratégie

Au terme de cet atelier, le groupe partage la conviction de l'importance de la mise en place d'un schéma directeur numérique (SDN) pour l'université cible⁴, afin de structurer la mise en œuvre de sa stratégie numérique.

L'enjeu d'un schéma directeur numérique est d'élaborer une stratégie globale intégrant les dimensions système d'information (SI), infrastructures, contenus numériques (les données !), services et usages.

Ce travail pourra être facilité par le fait que, selon les sujets, certains établissements seront leaders de pratique et pourront faire profiter de leur maturité numérique aux autres établissements.

La seconde partie du dyptique «Enjeux du numérique – Déluge de données», sur le thème « Numérique - organisation, process, outils » aura lieu à l'ESIEE Marne-la-Vallée, les 05 et 06/04/2018

Animateurs : Arnaud Bonnard (Ifsttar), Pascal Romon (UPEM)

Rapporteur : Didier Richard (ENSG)

⁴ La caisse des dépôts propose un guide afin d'aider les établissements à se mobiliser pour engager les transformations numériques indispensables <http://www.caissedesdepots.fr/guide-des-bonnes-pratiques-des-schemas-directeurs-numeriques-universitaires>