



HAL
open science

Les données de la recherche AAU-CRENAU : résultats de l'enquête sur les usages des chercheurs, doctorants et ingénieurs en matière de gestion de données

Laurence Bizien, Véronique Dom

► To cite this version:

Laurence Bizien, Véronique Dom. Les données de la recherche AAU-CRENAU : résultats de l'enquête sur les usages des chercheurs, doctorants et ingénieurs en matière de gestion de données. [Rapport de recherche] Centre de recherche nantais Architectures Urbanités. 2019. hal-02420916

HAL Id: hal-02420916

<https://hal.science/hal-02420916v1>

Submitted on 20 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Les données de la recherche AAU-CRENAU

Résultats de l'enquête sur les usages
des chercheurs, doctorants et ingénieurs
en matière de gestion de données

Laurence Bizien – Avril 2019

Collaboration éditoriale : Véronique Dom

Table des matières

Contexte : études de référence, public visé, objectifs, modalités.....	3
Synthèse	6
L'enquête statistique	7
Résultats	8
Perspectives et feuille de route	18
Annexe : le questionnaire.....	21

Études de référence

Hélène Prost, Joachim Schöpfel. Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. : Rapport final. [Rapport de recherche] Lille 3. 2015. <hal-01198379>

Alexandre Serres, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, Didier Collet. Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes [Rapport de recherche] Université Rennes 2. 2017, 159 p., 47 p., 114 p., 26 p., 23 p. <hal-01635186v2>

Julie Duprat. Les données de la recherche à l'Université Bordeaux Montaigne : Synthèse d'une enquête qualitative auprès des chercheurs. [Rapport de recherche] Université Bordeaux Montaigne. 2019. <hal-02020141>

Contexte

La mise en œuvre de cette enquête s'est décidée avec la directrice de l'équipe nantaise de l'UMR afin d'amorcer une réflexion au sein du CRENAU sur la gestion des données de la recherche et mettre en place un projet de politique dans le laboratoire ; cette même démarche était par ailleurs entreprise au CRESSON, par la diffusion d'un questionnaire réalisé par Léa Mosnier, apprentie au CNRS, qui devrait se poursuivre par des entretiens de chercheurs volontaires. En effet, conscient des enjeux liés aux données de la recherche, le laboratoire AAU réfléchit à mettre en place des services pour accompagner les chercheurs dans la gestion de leurs données. Les habitudes de travail sont actuellement isolées quand il s'agit de gérer celles-ci au sein du laboratoire.

Un questionnaire a ainsi été rédigé, très largement inspiré de celui des enquêtes effectuées dans les universités de Rennes et de Lille, et de la synthèse de l'enquête qualitative de l'université Bordeaux-Montaigne. Les résultats de l'enquête menée par le Groupe de travail gestion des données de la recherche de l'ULB, présentés aux documentalistes des laboratoires des écoles lors de leur réunion à Bruxelles en janvier 2019, ont complété cette documentation.

Public visé

Le public visé par cette enquête était celui des enseignants-chercheurs, ingénieurs de recherche et chercheurs du CRENAU, notamment les responsables du laboratoire, mais aussi les doctorants et post-doctorants. En effet, il semblait important de les associer, même si leurs besoins et pratiques peuvent s'avérer différentes, afin d'orienter les jeunes chercheurs dans cette démarche. Rappelons par ailleurs qu'un plan de gestion peut être associé à une thèse (ex : <https://riojournal.com/article/8664/>) , et que

ce plan commence à être imposé aux doctorants dans les pays anglo-saxons (ex. à l'Université de Bath : <https://library.bath.ac.uk/research-data/data-management-plans/university-dmp-templates>)...

Un séminaire de formation consacré à cette problématique et centré sur le traitement des données contenues dans les thèses de doctorat était ainsi organisé en 2017 à la Bibliothèque technique nationale de Prague : « Les thèses de doctorat représentent un bon terrain d'étude, à la fois négligé et à fort potentiel. » (<https://cefres.hypotheses.org/969>)

Objectifs

Les objectifs de cette enquête sont multiples :

- Établir un état des lieux : panorama sur les données de la recherche au CRENAU (types, volumes ...), connaître les pratiques des chercheurs mais aussi leur perception et réticences quant à l'ouverture et le partage de leurs données, définir les priorités dans les besoins (services, infrastructures, accompagnement)
- Sensibiliser les chercheurs : le questionnaire a été pensé comme une première étape permettant aux chercheurs une réflexion sur leurs pratiques et sur les enjeux et l'importance de cette thématique
- Identifier les chercheurs/doctorants sensibilisés à la problématique et pouvant être mobilisés autour d'une feuille de route projet ou outil
- Confronter les résultats à ceux obtenus par d'autres enquêtes, en particulier celle du CRESSON
- Légitimer des propositions d'offre de services répondant aux besoins exprimés, en terme d'accompagnement, de formation, de ressources et outils ...

Modalités

Les modalités de cette enquête ont été définies en faisant le choix d'une enquête simplement quantitative autour d'un questionnaire. Celui-ci a été validé par la directrice et transmis pour information en fichier PDF fin février, sa présentation étant à l'ordre du jour d'une réunion de laboratoire le 5 mars : cette prise de connaissance a ainsi permis un premier échange autour de la thématique. A l'issue de cette réunion, le questionnaire a été envoyé via GoogleForms à l'ensemble du CRENAU, avec un délai de réponse de 15 jours. Une relance a été faite dans la lettre d'information de l'UMR le 20 mars.

La restitution de cette enquête sera faite auprès de l'équipe avant le 20 mai, une demi-journée consacrée aux données de la recherche étant organisée dans le cadre des animations scientifiques du CRENAU. A la suite des résultats obtenus, des actions pourront ainsi être mises en place, la phase d'audit étant l'occasion de communiquer et de commencer à sensibiliser les chercheurs. La création d'une

politique de gestion des données est complexe, car il faut l'adapter à sa faisabilité au sein du laboratoire et trouver l'outil opérationnel adéquat. La mise en œuvre de cette politique est aussi liée à la capacité du laboratoire à réunir des compétences autour de ce projet (chercheurs, ingénieurs, informaticiens, personnels IST, archiviste, juriste, chargée de communication) au sein d'un groupe de travail.

En complément :

« Les chercheurs et leurs données en SHS » : retourner aux idéaux d'Humboldt ! (Journée d'étude du 15 juin 2018, MSHB, Rennes) in : <https://labedoc.hypotheses.org/2126>, Laurence Bizien, 25 juin 2018

« Les outils de gestion des données de la recherche dans les universités belges » in : <https://labedoc.hypotheses.org/5579>, 2 avril 2019

Atelier pratique destiné aux enseignants-chercheurs et doctorants de l'UMR AAU-CRENAU de l'école d'architecture de Nantes :

« Se former en jouant » in : <https://www.inist.fr/realisations/se-former-en-jouant/> Inist -CNRS, avril 2019

La synthèse

Un taux de participation très positif : 69.81 % !

Données sources	Les données d'observation (remontées de terrain, données de capteur) sont la source la plus importante (73%), suivie par les enquêtes et entretiens (70%) ; les données résultat : 73% les classent comme données textuelles, presque 60% des répondants les identifient comme graphiques, 40% statistiques, 32% 3D. Etait demandé de préciser la nature de ces données produites : les données textuelles occupent de loin le premier rang (75%)
Types de données	Quelques que soient les disciplines concernées, plusieurs types de données sont manipulées au CRENAU. Cette diversité pose la problématique du plan de conservation des données, avec des besoins de stockage adaptés. Le papier n'a pas disparu dans les pratiques, ce qui signifie que l'accès libre à certaines données produites n'est pas possible. Néanmoins, la mise en œuvre au sein du CRENAU du dépôt auprès des Archives départementales des documents produits (dossiers d'enquête, données travaillées, documentation scientifique, dossiers de travail, supports d'intervention ...) permettra la gestion et l'archivage de ces documents.
Stockage et sauvegarde	On constate que les répondants génèrent pour 45% entre 100Go et 1To par an. L'usage de périphérique de stockage externe est donc majoritaire, alors que celui-ci n'est pas très fiable. Le problème relevé est donc celui d'un risque de perte et d'un faible niveau de sécurité. Globalement, on constate une absence de politique de conservation.
Durée de conservation	Concernant la durée de conservation de leurs données, presque 95% des répondants la souhaite pérenne, à long terme et au-delà de la durée de leur projet, pour valider leurs résultats, mener de nouvelles études ou pour l'enseignement.
Accès	La position sur la diffusion de ses données de recherche est majoritairement positive, avec seulement 8% de réponses négatives catégoriques. « Ignorance plutôt que bonne ou mauvaise volonté » (cf. enquête de Lille). Ce positionnement manifeste des chercheurs, prêts pour déposer leurs résultats, est nuancé par des interrogations. Parmi les raisons qui inciteraient à rendre les données de recherche accessibles en libre accès, le souhait de voir se développer d'autres recherches à partir des jeux de données est la principale motivation évoquée (63%) ; Parmi les raisons pour ne pas déposer, en tête pour 64% des répondants, les freins juridiques (le caractère sensible ou confidentiel des données).
Types de dépôts	Préférences par rapport au dispositif : le support à privilégier pour un dépôt en libre accès semble être celui des archives ouvertes de type HAL pour 57%
Services attendus	La plupart des services proposés est demandée : en tête de la liste, des conseils d'ordre général pour la gestion des données (60%), puis des conseils techniques (normes, métadonnées etc.) et des conseils juridiques (57%). Les services de conseil spécialisés sont donc plébiscités, en collaboration avec des techniciens et informaticiens (métadonnées, système de nommage des fichiers ...).

L'enquête statistique

Le questionnaire

Les éléments et la structure des enquêtes précédentes ont été adaptés au périmètre de l'équipe CRENAU, selon un schéma directif de questionnaire comportant 27 questions avec des axes thématiques : champ disciplinaire/statut ; typologie des données ; pratiques de stockage et d'archivages ; pratiques de partage et de diffusion ; besoins et attentes. Le formulaire comportait une seule question ouverte afin d'en faciliter le dépouillement (question 15 concernant les données produites par les chercheurs non permanents), la restitution du questionnaire et la demi-journée consacrée à ce sujet le 20 mai devant permettre un espace d'échange. 11 questions étaient à choix unique ou fermé. 16 questions étaient à choix multiples et l'option Autre était proposée 4 fois, afin de pallier à la rigidité du sondage.

L'échantillon collecté

Le questionnaire a été envoyé à l'ensemble de l'équipe de chercheurs et doctorants du CRENAU, soit 23 chercheurs permanents et 6 contractuels (plus les chercheurs inscrits dans la liste de diffusion parmi les 9 chercheurs associés), et les 24 doctorants.

37 réponses ont été recueillies sur les 53 membres sollicités, soit **un taux de participation de 69.81 %**, donc un résultat très positif sachant que la moyenne du taux de réponse des enquêtes en ligne est de 20 à 30%. On peut donc considérer que ce taux de participation élevé constitue une image assez précise des pratiques au sein de l'équipe.

Pour information : 28% de participation à Rennes, 15% à Lille et à l'ULB.

Résultats

Le profil des répondants : la première partie du questionnaire (« champ disciplinaire et statut ») s'attachait à cerner le profil des répondants : la tranche d'âge, le statut, les responsabilités actuelles, les disciplines concernées.

Les statuts

Les doctorants et maîtres de conférences ont été les plus nombreux à répondre à l'enquête (20 et 12 répondants), suivis par les professeurs (4) et chercheurs (2), ingénieurs (1) et post-doctorants (1). Ce nombre de répondants doit être mis en rapport avec la répartition au sein de l'équipe pour obtenir la représentativité réelle des différentes catégories :

- Professeur : 4 sur 7 soit 57%
- Maître de conférence : 12 sur 17 soit 70%
- Chercheur : 3 sur 3 soit 100 %
- Doctorant : 20 sur 24 soit 83%
- Post-doctorant : 1 sur 1 soit 100%

A titre indicatif, le titre de HDR a été questionné : 2 HDR sur 7 ont répondu (28 %)

« Ce n'est pas la taille de l'effectif de départ qui influe le plus sur le taux de participation et le statut du répondant est un élément à prendre en compte tout comme les responsabilités » (Rennes)

L'âge des répondants

3 tranches d'âge avaient été proposées... La majorité a plus de 30 ans (23, 62%), suivi par les moins de 30 ans (10, 27%) et seulement 4 plus de 50 ans (10,8%).

Ces chiffres peuvent s'expliquer par le taux de réponse important des doctorants et la moyenne d'âge de l'équipe (40,8 ans en mars 2019).

L'âge moyen du corps des professeurs des écoles est de 56 ans, celui des maîtres de conférence 52 ans (Bilans de gestion 2018 MC)

Les responsabilités

Une question portait sur les responsabilités exercées des répondants, celles-ci étant susceptibles d'offrir des analyses croisées sur la problématique des données. 20 répondants ont rempli cette question. Les « 30/50 ans » déclarent avoir des responsabilités scientifiques (direction de laboratoire, d'équipe de recherche, de groupe projet, de projet ANR) pour 6 d'entre eux, 6 encadrent des thèses et 4 dirigent une revue ou sont membres d'un comité de rédaction ; les « plus de 50 ans » sont 2 à avoir des responsabilités scientifiques, d'encadrement de thèses et 1 éditoriales. Les « moins de 30 ans » sont 2 à mentionner des responsabilités éditoriales ou d'encadrement de projet. Il semble bien que la tranche 30-50 exerce le plus de responsabilités, et souvent plusieurs, au sein du CRENAU. Certains répondants

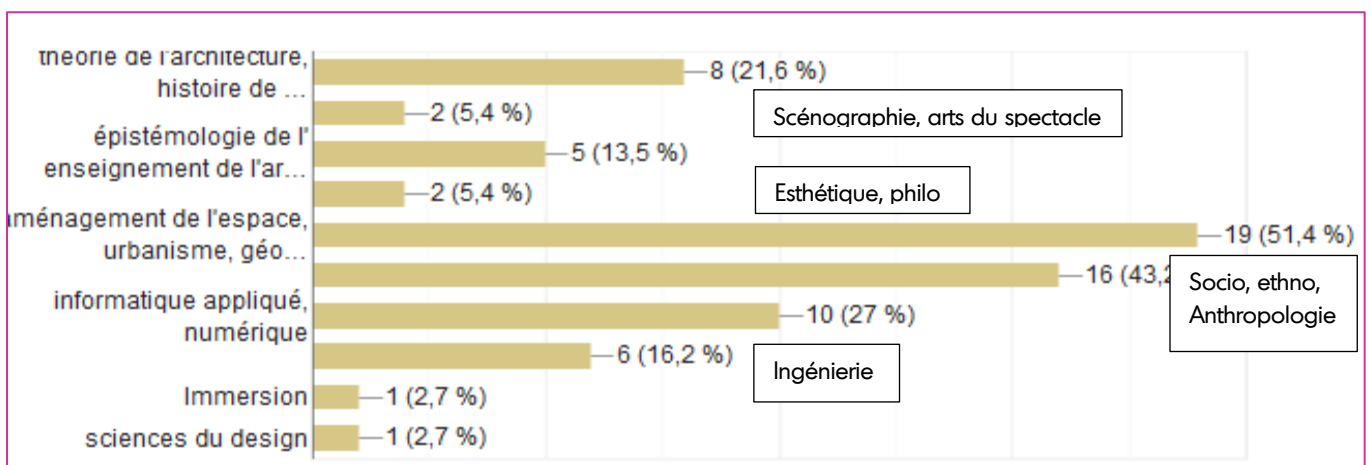
rappellent par ailleurs leurs responsabilités éducatives ou administratives non listées (encadrement de master, de stage, président du CPS).

Il était donc intéressant de croiser statut et responsabilités des répondants au sein du CRENAU, afin d'identifier les acteurs potentiels de cette dynamique, enclins à **s'approprier cette démarche individuelle** mais aussi, dans le cadre de leurs responsabilités, à **favoriser celle-ci** auprès des doctorants (HDR) et master de recherche, et se l'approprier de manière institutionnelle (direction). Les répondants actifs dans **l'édition scientifique**, déjà sensibilisés à l'open access, peuvent aussi poursuivre cette ouverture en l'élargissant à celle des données, renforçant ainsi **le partage des connaissances (open data, data paper) et l'engagement du laboratoire dans l'intégrité scientifique** (plagiat, transparence).

Les disciplines

L'ensemble des répondants ont renseigné la ou les discipline(s) scientifique(s) dont ils dépendent.

8 champs disciplinaires étaient proposés, en adéquation aux thématiques du laboratoire. Sur les 37 répondants, 19 ont coché « Aménagement de l'espace, urbanisme, géographie », 16 « sociologie, ethnologie, anthropologie », 10 « informatique appliqué et numérique », 8 « théorie de l'architecture, histoire de l'architecture », 6 « ingénierie » et 5 « épistémologie de l'enseignement de l'architecture » ; « scénographie » et « esthétique et philosophie » n'ont recueilli que 2 voix, et 2 répondants ont ajouté « Immersion » et « sciences du design » ...



Si l'on compare aux enquêtes multidisciplinaires des universités Rennes 2 et Lille 3, on constate qu'à Rennes les chercheurs en « Architecture et arts appliqués » représentaient 9% des répondants, ceux de l'« aménagement de l'espace, urbanisme » 3%, « géographie » 6%, « sociologie » 7%, « informatique » 1%; Lille a recueilli 5 réponses sur 270 en « Architecture et arts appliqués », 1 en « géographie ». La participation dépassait les 10% pour les 7 facultés de l'ULB, dont 12% pour la faculté d'Architecture la Cambre-Horta.

Cela donne une image intéressante des disciplines concernées et intéressées, mais seul un rapprochement avec les résultats de la faculté d'architecture semble pertinent, la comparaison avec les enquêtes universitaires où l'architecture n'est pas enseignée en tant que telle et où l'enseignement artistique souvent une simple composante s'avérant difficile.

La question d'un changement de pratiques plus culturel que technologique est néanmoins posée (cf formation INRA), par exemple par Marta Teperek and Alastair Dunning : <https://blogs.lse.ac.uk/impactofsocialsciences/2018/11/14/the-main-obstacles-to-better-research-data-management-and-sharing-are-cultural-but-change-is-in-our-hands/>



© 2019 Springer Nature Publishing AG

La typologie des données (collectées et produites)

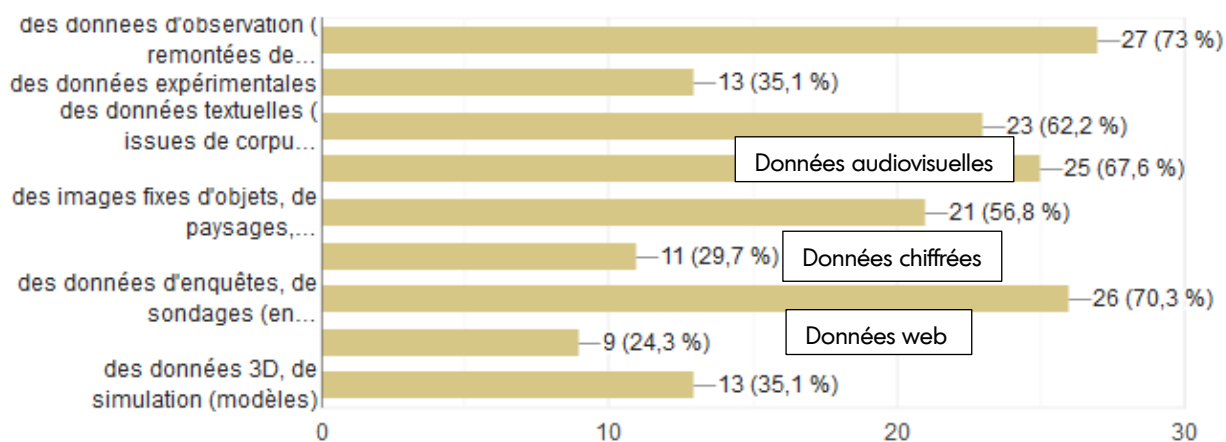
Quelles sont les données de la recherche que les chercheurs produisent comme résultat de leur travail scientifique au sein du CRENAU ? Les questions portaient sur les données sources, utilisées par les chercheurs pour faire leur travail et d'autres sur les données résultats, produites par les chercheurs pendant et à l'issue de leur travail. Les catégories (typologie), issues des enquêtes SHS, ont été adaptées au laboratoire (données 3D, simulation). Plusieurs réponses étaient possibles.

L'ensemble des participants a répondu à la question des **données sources**, les **données d'observation (remontées de terrain, données de capteur)** sont la source la plus importante (73%), suivie par les **enquêtes et entretiens** (70%), les données audiovisuelles (67%), les données textuelles (corpus, archives) pour 62% et les images pour 56%. Moins importants : les données expérimentales et 3D (35%), les données chiffrées (29%), et finalement les données web (24%).

A la question sur le **mode d'obtention de ces données**, a été répondu que 71% proviennent du laboratoire (production interne des données), 40% d'un tiers (données produites dans le cadre d'un projet collaboratif ou récupération de données exposées), 25% d'une base de données (sauvegarde, téléchargement d'un corpus issu d'une base externe).

Dans quelles catégories suivantes classeriez-vous vos "données sources" (vos données collectées)? Plusieurs réponses possibles

37 réponses



Concernant le **support matériel** des données sources, 52% ont déclaré qu'elles sont nativement numériques, 41% numérisées, 22% non numériques. Seules les personnes ayant répondu « Non numériques » à la question pouvaient répondre à la suivante concernant la numérisation de celles-ci (20 répondants) : 80% a répondu oui.

Les questions suivantes portaient sur les **données résultat**, avec une proposition de classement par catégories après exploitation.

73% les classent comme données textuelles, presque 60% des répondants les identifient comme graphiques, 40% statistiques, 32% 3D. Était demandé de préciser la nature de ces données produites : **les données textuelles occupent de loin le premier rang (75%)**, puis les photographies (48%), les enregistrements audio (40%), les tableaux et les visualisations (32%), des données spécifiques (29%), des bases de données et des enregistrements vidéo (27%). Les programmes et applications arrivent en dernier (24%). 32 répondants ont détaillé leurs réponses : transcriptions d'entretiens pour 75%, cartes et plans pour 50%, annotations et corpus de texte pour 43%, analyses statistiques (40%), questionnaires (34%), série de mesures (25%). Les cartographies recueillent 15%, les images satellites 9%.

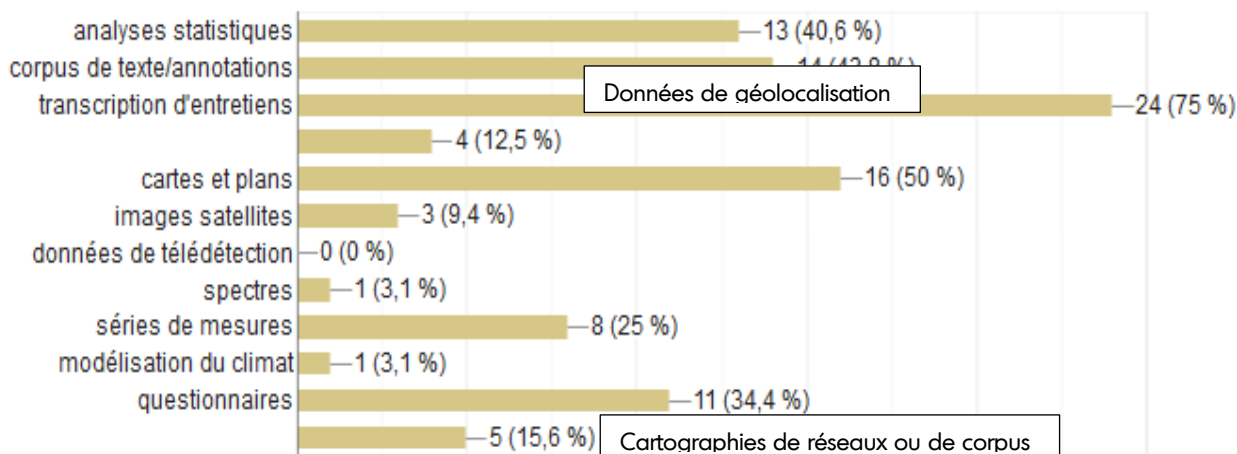
La diversité des réponses reflète ainsi celle des données produites.

Quelques que soient les disciplines concernées, plusieurs types de données sont manipulées au CRENAU. **Cette diversité pose la problématique du plan de conservation des données, avec des besoins de stockage adaptés. Le papier n'a pas disparu** dans les pratiques, ce qui signifie que l'accès libre à certaines données produites n'est pas possible. Néanmoins, la mise en œuvre au sein du CRENAU du **dépôt auprès des Archives départementales** des documents produits (dossiers d'enquête, données travaillées, documentation scientifique, dossiers de travail, supports d'intervention ...) permettra la gestion et l'archivage de ces documents.

(Cf. Référentiel de gestion des archives de la recherche – AAF section Aurore)

Veillez détailler :

32 réponses



Pratiques de stockage et d'archivages des données

On constate que les répondants génèrent pour **45% entre 100Go et 1To par an**, et entre 1 et 20 Go pour 37%. 13% pensent produire plus d'1 To. En comparant avec les résultats de la faculté d'architecture de La Cambre-Horta, on constate presque les mêmes pourcentages : 45% entre 100 Gb et 1 Tb, 22% entre 50 Gb et 100 Gb.

78% des participants ont indiqué comme espace principal de stockage **un disque externe** (ou autre support) ; la majorité privilégie le stockage local, sur leur ordinateur personnel (56%) ou professionnel (54%). 51% stockent « dans le cloud », 18% utilisent le serveur du CRENAU, alors que 13% ont des données sur le serveur d'une autre institution (8% dans le serveur d'une TGIR).

L'usage de périphérique de stockage externe est donc majoritaire, alors que celui-ci n'est pas très fiable. Le problème relevé est donc celui d'un risque de perte et d'un faible niveau de sécurité.

Concernant la **fréquence de sauvegarde des données**, seuls 16% déclarent sauvegarder quotidiennement leurs données, alors que 30% le font d'une manière irrégulière « en fonction des besoins » et 8% tous les ans. D'autres sauvegardent plus souvent, une fois par semaine (8%), par mois (27%) ou par trimestre (8%). La question peut cependant révéler des interprétations différentes quant au terme « sauvegarde » : enregistrement des mises à jour, dont la sauvegarde est quotidienne, archivage des dossiers à fréquence plus espacée. (Cf. enquête de Lille)

Globalement, on constate une absence de politique de conservation.

Quant à la **durée de conservation de leurs données**, presque **95% des répondants la souhaite pérenne, à long terme** et au-delà de la durée de leur projet, pour valider leurs résultats, mener de nouvelles études ou pour l'enseignement. Pour ce qui est de leur responsabilité sur la sauvegarde des données et leur archivage, 35% déclarent avoir mis au point **une politique d'archivage** personnelle, en l'absence d'une politique du laboratoire (critères, modalités, nommage de fichiers...).

La question suivante portait sur la **conservation des données de chercheurs non-permanents** (doctorants, chercheurs associés, partenaires...) : cette question était optionnelle et libre.

17 réponses ont été recueillies, les doctorants mentionnant le disque dur externe (1 To) leur étant distribué. Une majorité indique que ces données devraient être conservées au laboratoire, comme celles des membres permanents, à long terme, comme livrable obligatoire.

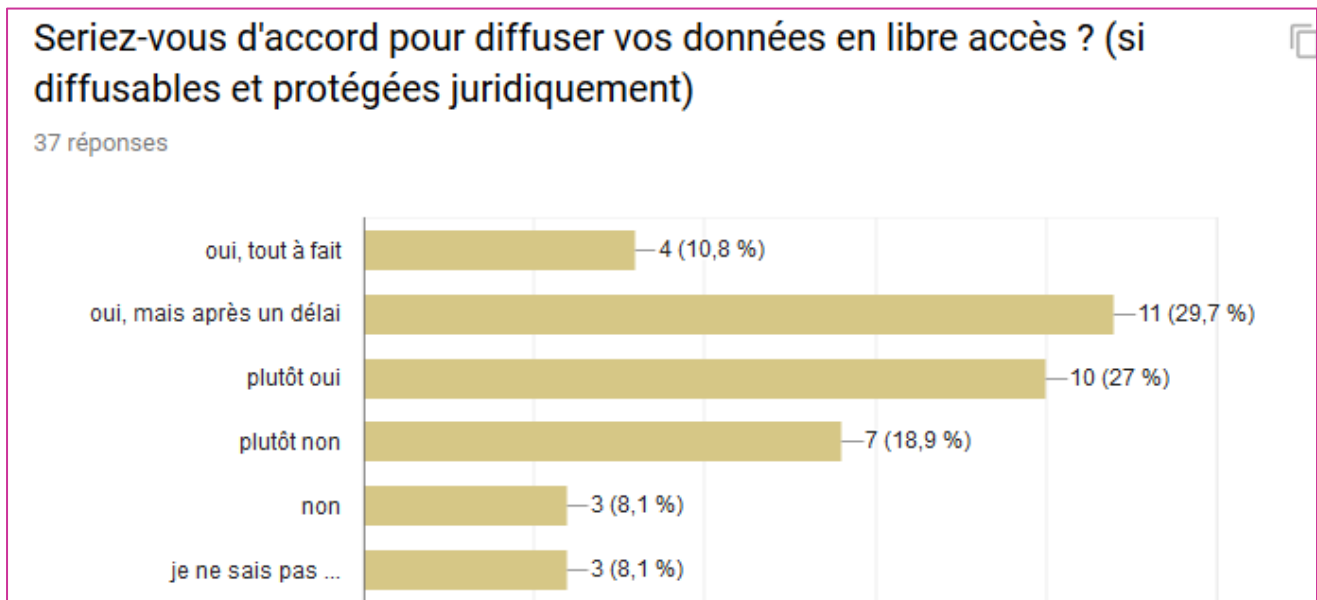
Le risque de perte des données lors du départ des chercheurs ou doctorants est identifié. Le serveur du laboratoire est privilégié.

Pratiques de partage et de diffusion des données

A la question « qui, en dehors de vous, peut accéder à vos données? », **70% des participants ont répondu : « personne d'autre »** ; l'accès est ensuite autorisé aux collègues du projet de recherche, du groupe de travail (30%). En dehors de l'équipe scientifique, au sein de l'unité ou dans le cadre d'un projet, **les données du CRENAU sont donc largement inaccessibles ...** La solution de partage doit permettre un accès aux partenaires du groupe et/ou projet, éventuellement d'autres institutions.

Sur la pratique du libre accès et d'un éventuel **dépôt de données dans une archive ouverte** type Zenodo, aucun n'a encore pratiqué un tel dépôt, par méconnaissance de cette possibilité (46%). D'autres ne savent pas comment (8%), n'ont pas le temps (24%), 13% seraient prêts à le faire avec un soutien, et 13% dit clairement qu'ils n'ont pas l'intention de partager leurs données via le web.

La position sur la diffusion de ses données de recherche est majoritairement positive, avec seulement 8% de réponses négatives catégoriques. « Ignorance plutôt que bonne ou mauvaise volonté » (cf. enquête de Lille). Ce positionnement manifeste des chercheurs, prêts pour déposer leurs résultats, est nuancé par des interrogations.

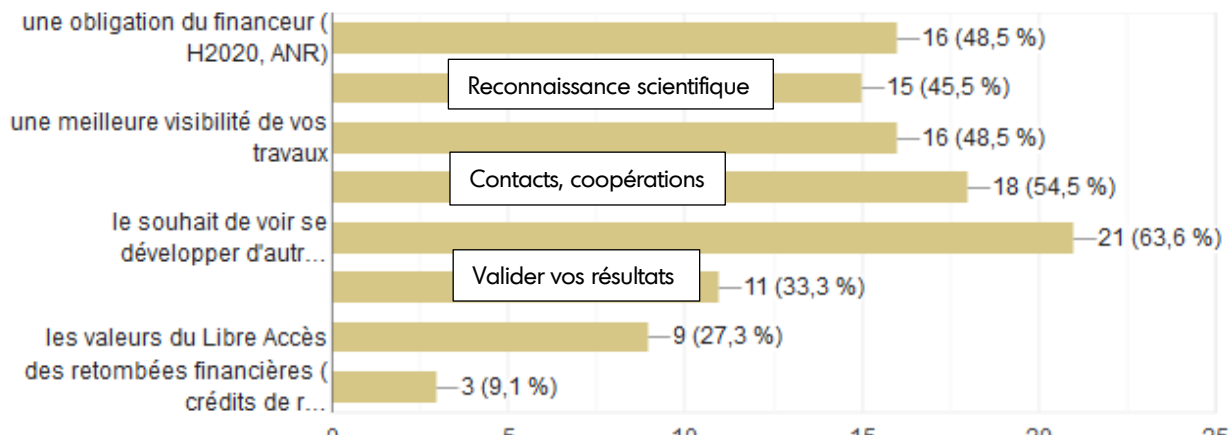


Parmi **les raisons qui inciteraient** à rendre les données de recherche accessibles en libre accès, le souhait de **voir se développer d'autres recherches à partir des jeux de données est la principale motivation évoquée (63%)**, puis celle d'être l'occasion de nouveaux contacts, de nouvelles coopérations scientifiques (54%) ; Une meilleure visibilité des travaux de recherche et une obligation faite par le financeur obtiennent 48%, la reconnaissance de la communauté scientifique 45%, le souci de pouvoir

faire valider les résultats à partir des jeux de données 33% ; l'adhésion aux valeurs de Libre Accès aux résultats de la recherche arrive en 6^e position (27%), juste avant de possible retombées financières 9%.

Parmi ces raisons, lesquelles vous inciteraient à rendre vos données accessibles en libre accès ?

33 réponses

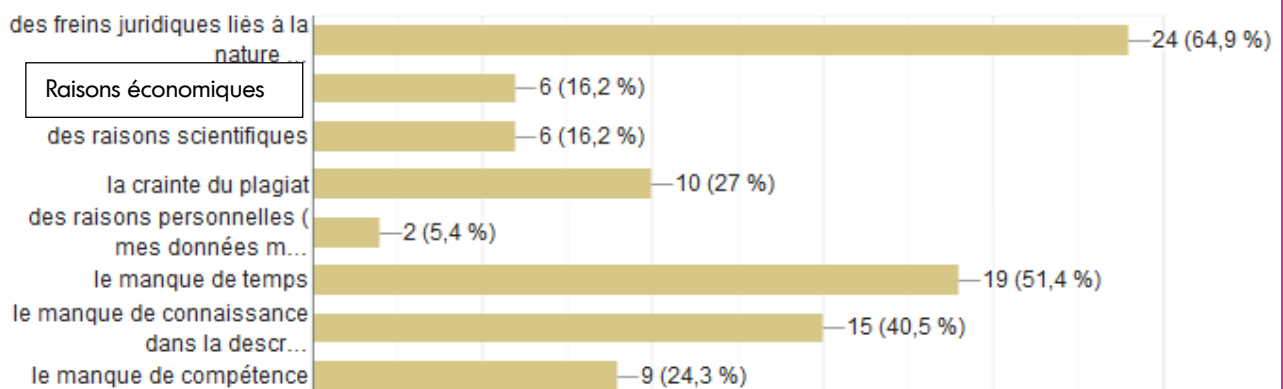


Parmi les raisons pour ne pas déposer, en tête pour 64% des répondants, les freins juridiques (le caractère sensible ou confidentiel des données), puis le manque de temps pour 51% et le manque de connaissance (description des jeux de données, modalités de dépôt) à 40% ou de compétence (24%).

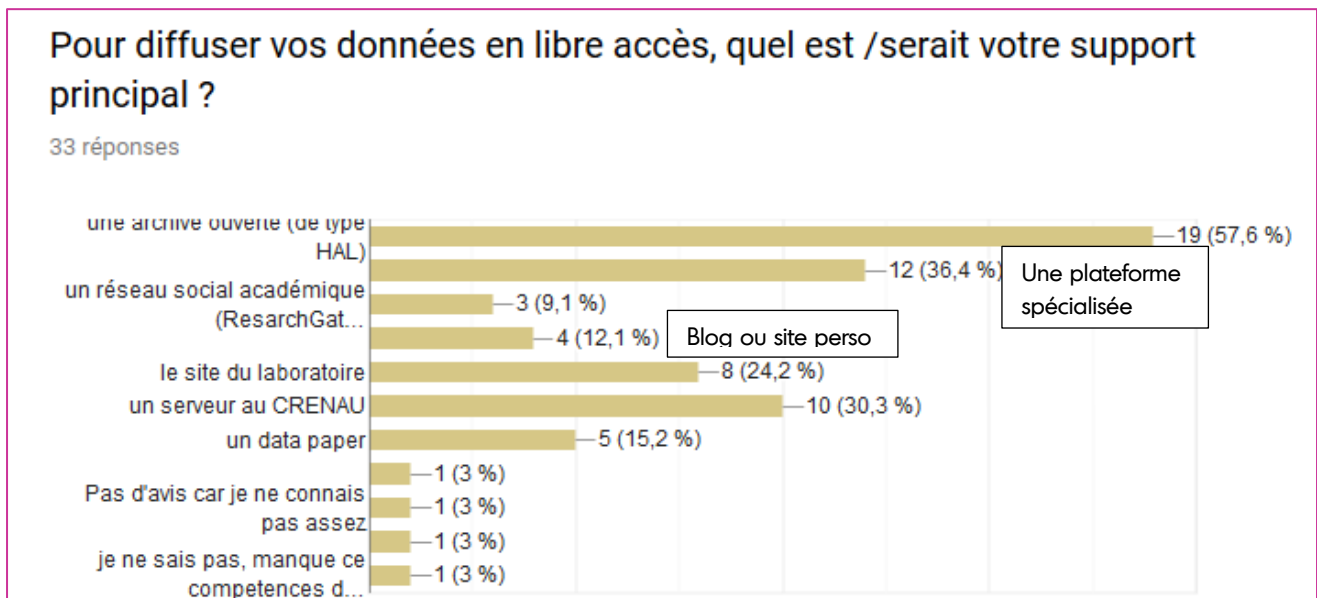
D'autres raisons évoquées sont la crainte du plagiat (27%), des raisons économiques ou scientifiques (16%) ou plus simplement la protection de leur propriété intellectuelle: « ces données m'appartiennent » ... (5%).

Quelles seraient les raisons qui vous empêcheraient ou vous freineraient pour rendre vos données de recherche accessibles en libre accès ?

37 réponses



Préférences par rapport au dispositif : le support à privilégier pour un dépôt en libre accès semble être celui **des archives ouvertes** de type HAL pour 57% ; ensuite une plateforme spécialisée dans l'archivage pour 36%, le serveur local du CRENAU pour 30%, puis le site du laboratoire (24%).



Un dépôt dans une infrastructure institutionnelle est donc retenu, même si aucun n'a réellement déjà utilisé ces dispositifs, mais le fait d'avoir déjà pratiqué des dépôts dans HAL influe sans doute la réponse. Cette solution est en correspondance avec la nécessité d'un archivage et d'un espace de travail sécurisé demandé en offre de service (54%).

Besoins et attentes

Pour finir, plusieurs questions étaient posées pour cerner les besoins des membres du laboratoire, en proposant une liste de services potentiellement intéressants ...

D'abord, pour connaître leurs connaissances et positions sur les données de recherche, une question portait sur leur connaissance d'un plan de gestion de données et des TGIR : **70% ne savent pas ce qu'est un plan, 78% une Très Grande Infrastructure de recherche...**

Une enquête réalisée en 2017 auprès d'un échantillonnage de chercheurs européens montrait encore de la méconnaissance (un quart des répondants ne savait pas ce qu'était un DMP non plus) :

https://ec.europa.eu/research/openscience/pdf/os_skills_wgreport.pdf

Leur position personnelle faisait l'objet de 6 questions fermées :

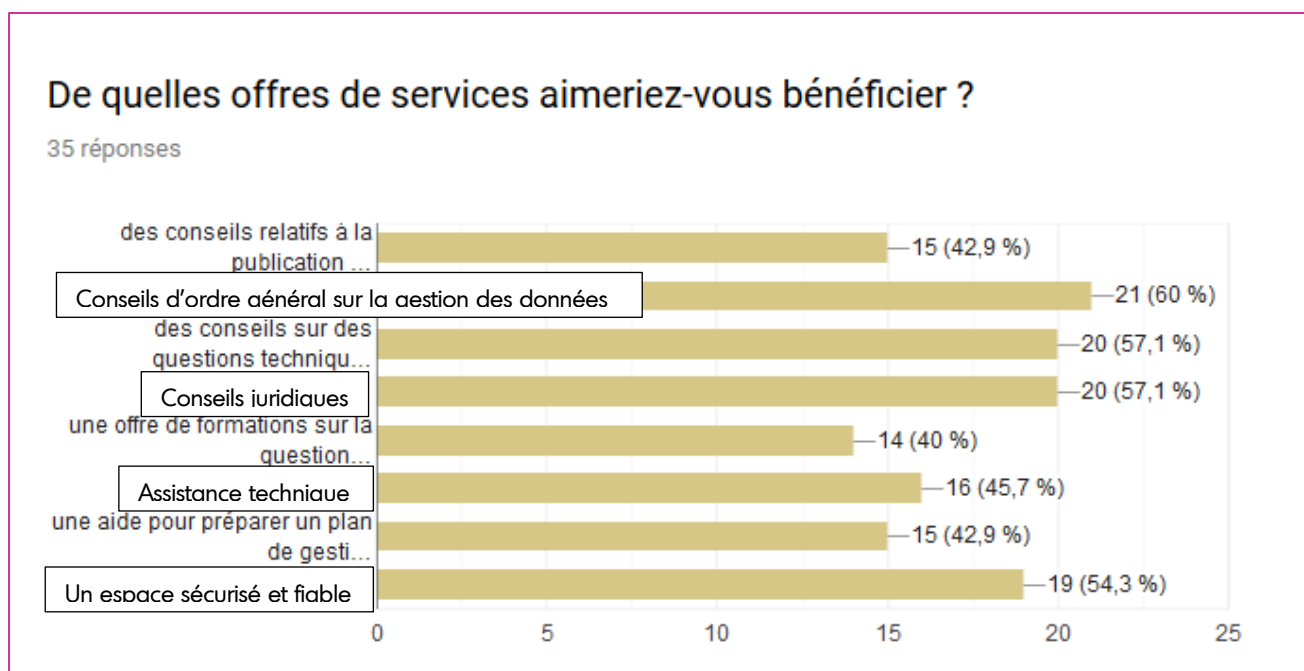
Je suis familier avec les enjeux de l'open access (accès ouvert)	oui : 51%
Je suis prêt à partager mes données avec les autres	oui : 83%

Je souhaiterais pouvoir stocker mes données au-delà de la durée du projet de recherche	oui : 97%
Je souhaiterais que le CRENAU dispose d'un plan de gestion de données (33 réponses)	oui : 91%
Je souhaiterais que le CRENAU dispose d'un ensemble de métadonnées prescrit pour verser mes données vers un entrepôt (29 réponses)	oui : 76%
Je souhaiterais que le CRENAU utilise un système de nommage standard de fichiers (30 réponses)	oui : 66%

On constate que certaines questions n'ont pas été renseignées par tous les répondants, sans doute faute de bien connaître les enjeux.

Une liste ouverte de huit services potentiellement intéressants pour les doctorants, enseignants-chercheurs etc.

Plusieurs réponses ont été possibles :



Selon les réponses, la plupart des services proposés sont demandés. **En tête de la liste, des conseils d'ordre général pour la gestion des données (60%), puis des conseils techniques (normes, métadonnées etc.) et des conseils juridiques (57%).**

L'espace d'archivage sécurisé et fiable pour ses propres données obtient 54% des demandes ; une assistance technique est souhaitée pour 45%, une aide pour préparer un plan de gestion et des

conseils relatifs à la publication et citations des données pour 42%. Bien que les chercheurs interrogés déclarent ne pas connaître ce dispositif à 70%, ils ne sont pas contre cette démarche et souhaitent donc être accompagnés. Certains chercheurs savent qu'ils vont être confrontés à cette problématique, puisque le DMP (Data Management Plan) est de plus en plus exigé par les bailleurs de fonds (ANR 2019).

En dehors du DMP, certains sont également sensibles à la diffusion de celles-ci et leur publication. Des liens entre les données et les publications doivent en effet être pensés dans le cadre du principe d'interopérabilité.

Grâce au DOI de la ou des publications qui exploitent un dataset et à celui du dataset lui-même, un lien bidirectionnel pérenne est ainsi créé ; il est par exemple exploité dans ScienceDirect avec l'entrepôt Pangaea et dans Scopus dans l'encart Related Research Data.

De plus, certains éditeurs et financeurs exigent maintenant la publication des données qui accompagnent notamment une publication : des revues ont désormais une politique de données qu'on trouve dans leurs pages Data Policy (cela concerne les données qui pourraient servir de preuve à la démonstration scientifique présentée dans un article).

Certains éditeurs préconisent voire imposent un ou plusieurs entrepôts pour y déposer les données (exemple : groupe Springer Nature, Geoscience Data Journal). Certains éditeurs encouragent donc la publication des données accompagnant l'article, mais il faut veiller aux conditions d'accès et de réutilisation de celles-ci, au regard des exigences d'une Science Ouverte. Il faut aussi penser à leur usage en fonction de la protection des données à caractère personnel (RGPD). Depuis décembre 2018, l'ajout d'un champ duplicable « données associées » permet ainsi d'ajouter dans la notice HAL le DOI fourni par l'entrepôt dans lequel sont archivées et donc les données associées à la publication.

Enfin, pour une meilleure valorisation d'un jeu de données, il est possible de publier un data paper (ou software paper pour un logiciel). En dernier, moins nombreux sont ceux qui aimeraient une offre de formations (40%). **Les services de conseil spécialisés sont donc plébiscités, en collaboration avec des techniciens et informaticiens (métadonnées, système de nommage des fichiers ...).**

Cette enquête statistique comportait deux questions finales sur le souhait de bénéficier d'une présentation des plateformes Humanum-Loire et Progedo (MSH Guépin), et d'un atelier sur la constitution d'un plan de gestion de données. **93% et 90% des répondants se déclarent intéressés par ces deux propositions.**

En raison du départ de la correspondante Humanum en juin, cette présentation devait être programmée au plus tard fin mai. La date du 20 mai a été retenue, avec une diffusion au préalable auprès des répondants des résultats de ce sondage.

Perspectives et feuille de route

Cette enquête a été conçue comme un outil de communication et de sensibilisation, pour susciter l'attention des personnels scientifiques du CRENAU et préparer la mise en place d'une démarche au sein de celui-ci. **Le fort taux de participation laisse à penser qu'une telle démarche peut être mise en place progressivement**, notamment en s'appuyant sur un groupe de chercheurs (et doctorants) plus avancés en matière de bonnes pratiques et de modèle, qui motiverait les autres. Une forme de méconnaissance sur les solutions possibles pour cet archivage est observée, complexifiée par un contexte juridique en évolution (Loi pour une république numérique, réglementation sur la protection des données personnelles). Faute de moyens, cette enquête ne sera pas suivie d'entretiens permettant d'approfondir celle-ci, contrairement à celle menée au CRESSON.

Les axes de cette politique s'inscrivent ainsi dans un contexte favorable, mais cette dynamique doit fédérer et mobiliser plusieurs acteurs (direction, informaticiens, professionnels de l'IST, archiviste, service juridique, éditeurs) pour déployer une ligne directrice, une feuille de route associant formation, assistance et conseil, infrastructure.

La rédaction de plans de gestion permettrait une réflexion en interne sur les processus de sauvegarde, de transparence et de partage à mettre en place, la pertinence de modèles propres au laboratoire (un template ANR est attendu en juillet), et la valorisation de ces données ...

Le rôle et les responsabilités de chaque acteur doivent donc être définis. Le CRENAU peut aussi s'appuyer sur différents outils externes, comme les plateformes Humanum et Progedo.

La documentaliste peut aussi exploiter l'outil DorANUM (plateforme d'auto-formation à la gestion des données) en complément d'une formation en présentiel, selon le profil des chercheurs.

Huma-Num semble particulièrement intéressant car cette structure possède plusieurs outils (ShareDocs, HumaNum Box, Nakala) assurant la pérennité des données dont elle a la charge. Pour partager les données, l'entrepôt MediHAL pour les enregistrements oraux et leurs transcriptions (moissonné par Isidore) ou encore l'infrastructure européenne Zenodo (stockage à hauteur de 500 Go par jeux de données) sont des pistes à envisager.

Propositions des responsabilités des différents acteurs

Le rôle de la direction sera de piloter ce chantier et de mutualiser outils et compétences.

- Le rôle de coordinateur avec les acteurs du CRENAU et de l'UMR, mais aussi avec des acteurs externes est essentiel (École doctorale, MSH Guépin, Inist, archives départementales).

- Un recensement des acteurs impliqués et **la création d'un groupe de travail de volontaires** (enseignants-chercheurs, informaticien, documentaliste, archiviste, chargée de communication...) peuvent être associés à la mise en place d'actions de sensibilisation, de formations, des animations thématiques, l'encouragement à l'élaboration de DMP CRENAU, de documents d'accompagnement sur ce sujet.

Le rôle de la documentaliste au sein du laboratoire pourra être celui de :

- conseiller sur les questions de droit d'auteur, de législation autour des données ;
- apporter un soutien pour la rédaction du plan de gestion ;
- continuer à sensibiliser les chercheurs au libre-accès ;
- aider à la valorisation du projet (signalement de corpus) ;
- être référent pour la conservation des données non-numériques (archives) ;
- faire le lien avec le délégué à la protection des données de l'école ;
- accompagner une réflexion sur l'éthique et l'intégrité scientifique ;
- proposer des stages dédiés (INIST-CNRS, URFIST), en lien avec la correspondante formation de l'unité.

Le rôle de l'informaticien sera celui de :

- recommander une solution d'hébergement ;
- accompagner les agents sur les infrastructures choisies, avec des solutions de stockage destinées à garantir l'accès et la préservation des données à long terme et permettant la citation et le lien avec les publications.

Feuille de route proposée

Étape 1 : Mars –avril 2019 Sensibilisation

Audit sur les pratiques et besoins, actions de sensibilisation au sein du laboratoire (débat lors de réunions, présentation du serveur local et sensibilisation par l'informaticien de l'outil de stockage et de partage qu'est le serveur CRENAU, avec sauvegardes régulières.

Afin d'acquérir des compétences dans ce domaine et répondre aux besoins immédiats de création de DMP, la documentaliste suivra une formation à l'URFIST Rennes le 26 avril « Pourquoi et comment rédiger un plan de gestion de donnée »).

début de rédaction de « DMP pilotes » version test autour de 2 projets ANR coordonnés par des chercheurs CRENAU (Coolscares et Percilum)

Étape 2 : Mai - octobre 2019 Mobiliser et accompagner les chercheurs et les personnels

Animation autour du DMP et données (20 mai) avec Développement de solutions de stockage et de partage des données , politique d'information sur les entrepôts de données SHS (HumaNum, Nakala, Progedo) ; aider les chercheurs dans l'élaboration de DMP pour les projets de recherche (ANR) en cours ; création d'un groupe de travail au sein du CRENAU et/ou de l'UMR ; comparer les résultats des deux enquêtes UMR ; Sensibilisation des chercheurs à la question d'archivage des données sous format papier par la documentaliste et la référente archive de l'école ; établir une liste de recommandations disponibles dans un espace du serveur

Étape 3 : Novembre 2019 – 2020 Formaliser

Mise en place d'animations autour des données au sein du laboratoire (demi-journées sur les données sensibles, l'intégrité scientifique, la rédaction de data paper ...) ; proposer des outils d'enregistrement et de traçabilité des données (fiches projets, rapports) , réfléchir à un schéma et aux standards/vocabulaires des métadonnées (vocabulaire ArchiRès/Ambiance ; DARIAH ; ICOS ?) ; étudier la pertinence de créer un ou plusieurs modèles de Plan de gestion de données propre au laboratoire (modèle Opidor) : https://dmp.opidor.fr/public_templates ; Prévoir un cahier des charges concernant le stockage des disques dur distribués aux doctorants (cf infra) ; organiser un espace

collaboratif avec les partenaires avec la mise en place d'un partenariat avec la TGIR humanum-Num : sauvegarde des données dans un serveur collectif, dépôt des données « froides » dans Nakala ; réalisation d'un guide de rédaction d'un DMP

(http://www.bu.parisdescartes.fr/doc/recherche/Realiser_un_DMP_V2_2018_Def_HRS4R.pdf) ou d'une trame en format PDF comme à l'INRA (<https://www6.inra.fr/datapartage/Generer/Rediger-un-plan-de-gestion>), liste de recommandations, bibliographie et webographie, outils comme à l'EPFL (<https://researchdata.epfl.ch/plan-fund/dmp/>)

Étape 4 : 2020 Rendre visible

Rendre visible la politique du laboratoire avec :

- des pages sur le site internet sur la politique open data du laboratoire,
- le lien vers l'entrepôt choisi ,
- la rédaction d'un espace conseil avec guides juridiques et recommandations sur le site internet ou dans l'espace partagé (ex. <https://espacechercheurs.enpc.fr/fr/node/1094>).

Exposer et valoriser les données et publications (lien vers les données exposées, moissonnage Isidore) ;

Publier des data papers (publication qui décrit un jeu de données scientifiques) dans un data journal (revue dédiée à ce type de publication) ou dans une revue classique

Annexe : le questionnaire

Questionnaire : enquête sur les pratiques en matière de données de la recherche Mars 2019

Bonjour,

Merci de bien vouloir remplir ce questionnaire qui vise à mieux connaître vos pratiques et besoins en matière de données de la recherche ; il permettra d'établir un état des lieux de la situation des données au CRENAU et de réfléchir à une offre de services.

Les données deviennent un enjeu majeur pour la gestion et la diffusion de la recherche scientifique. De plus, tant pour les projets européens (Horizon2020) que nationaux (ANR), il faut fournir un plan de gestion des données.

L'enquête est anonyme, les résultats permettront de mettre en place une politique sur les données de la recherche et de vous faire des propositions.

Ce questionnaire est largement inspiré d'une enquête réalisée par l'URFIST de Rennes, la MSHB et le SCD Rennes 2 et de celle de l'Université de Lille 3.

CHAMP DISCIPLINAIRE ET STATUT

1. Vous êtes :

- Professeur
- Maître de conférences
- Ingénieur
- Doctorant
- HDR
- Chercheur

2. Votre tranche d'âge

- Moins de 30 ans
- 30-50 ans
- Plus de 50 ans

3. Exercer vous actuellement des responsabilités ?

- Scientifiques : direction de laboratoire, d'équipe de recherche, de projet ANR ...
- D'encadrement de thèses
- Editoriales (direction de revue, membre d'un comité de rédaction ...)
- Autre responsabilité :

4. Votre/vos discipline(s) :

- théorie de l'architecture, histoire de l'architecture
- scénographie, arts du spectacle, muséologie
- épistémologie de l'enseignement architecture
- esthétique, philosophie

- aménagement de l'espace, urbanisme, géographie
- sociologie, ethnologie, anthropologie
- informatique appliqué, numérique
- ingénierie

TYPOLOGIE DES DONNÉES

5. Dans quelles catégories suivantes classeriez-vous vos « données sources » (vos données collectées) ? (plusieurs réponses possibles)

- Des données d'observation (remontées de terrain, données de capteurs ..)
- Des données expérimentales
- Des données textuelles (issues de corpus de textes, d'archives...)
- Des données audiovisuelles (enregistrements audio, séquences filmiques ...)
- Des images fixes d'objets, de paysages, d'architectures, scans de textes
- Des données chiffrées (tableaux de données statistiques ...)
- Des données d'enquêtes, de sondages (entretiens)
- Des données web (données d'usage, cartographies de sites, données de réseaux sociaux ...)
- Des données 3D, de simulation (modèles)
- Autre :

6. Mode d'obtention des données

- Données générées par le laboratoire
- Données produites par un tiers
- o Contenus d'une BDD (base de données)

7. Concernant leur support matériel, vos données sources (ou données collectées) sont-elles majoritairement : UNE SEULE RÉPONSE !

- non numérique (par ex. : sources archivistiques imprimées, entretiens non enregistrés en numérique ...)
- Merci d'aller à la question 8*
- numérisées (par ex. : corpus numériques, des données statistiques)
- nativement numériques (par ex. : données web, entretiens numériques)

8. Si vos données sont non numériques, est-ce que vous les numérisez en général (saisie d'entretiens, numérisation des textes ...) ? UNE SEULE REPONSE !

- OUI
- NON

9. Dans quelles grandes catégories classeriez-vous vos données produites, après exploitation (vos données résultat) ? (plusieurs réponses possibles)

- Des données statistiques (données chiffrées)
- Des données textuelles (extraits, champs textuels, extraits d'entretiens ...)
- Des données graphiques (schémas, visualisations, cartographies, tableaux, scans de textes ...)
- Des données audiovisuelles (film ...)
- Des données 3D, de simulation
- Autre : ...

10. Pouvez-vous préciser la nature des données produites à l'issue de vos recherches ?

- photos
- visualisations ou modèles multidimensionnels
- enregistrements audio
- enregistrements vidéo
- textes
- tableaux
- bases de données
- programmes et applications
- autres, données spécialisées ou spécifiques : *merci de remplir la question 11 :*

11. Veuillez détailler :

- analyses statistiques
- corpus de texte/annotations
- transcription d'entretiens
- données de géolocalisation (SIG)
- cartes et plans
- images satellite
- données de télédétection
- spectres
 - séries de mesures
 - modélisations du climat
 - questionnaires
 - cartographies de réseaux ou de corpus

PRATIQUES DE STOCKAGE ET D'ARCHIVAGE DES DONNEES

12. A combien estimez-vous l'espace nécessaire au stockage de vos données ? **UNE SEULE REPONSE !**

- Moins d'1 Go
- Entre 1 et 20 Go
- Entre 100 Go et 1 To
- Plus d'1 To
- Ne sais pas, pas d'avis

(Pour information : clé USB de 1 à 20 Go ; disque dur autour de 1 To, serveur ou cloud au-delà du To)

13. Où stockez-vous les données de la recherche ? (*plusieurs réponses possibles*)

- sur votre ordinateur personnel
- sur votre ordinateur professionnel
- sur un disque dur externe ou autre support
- sur un serveur de l'école/labo
- sur une plate-forme privée dans le cloud (Google drive, Dropbox ...)
- sur un serveur externe d'une autre institution
- sur le serveur d'une TGIR (Très Grande Infrastructure de Recherche) (par ex. Huma-Num, progedo)

13. A quelle fréquence faites-vous ces copies de sauvegarde ? *UNE SEULE REPONSE !*

- tous les jours
- toutes les semaines
- tous les mois
- tous les trimestres
- tous les ans
- irrégulièrement, en fonction des besoins
- en fonction du prestataire de-je ne fais pas de copie ...
- je ne fais pas de copie ...

14. Combien de temps les données sont-elles /devraient être conservées et préservées ?

- Durée du projet
- long terme au-delà du projet (pour valider vos résultats de recherche, pour mener de nouvelles études ou pour l'enseignement)

14. avez-vous mis au point vous-même une politique d'archivage de vos données ?

- Oui
- Non

15. Concernant les données produites par des chercheurs non-permanents du laboratoire (doctorants, chercheurs associés, partenaires), comment le laboratoire assure-t-il la conservation et la pérennisation des données ?

PRATIQUE DE PARTAGE ET DE DIFFUSION DES DONNEES

16. En dehors de vous, qui peut accéder directement à vos données ?

- personne
- mon groupe de travail/recherche
- le laboratoire
- tout le monde !
- autre :

17. Avez-vous déjà stocké vos données en libre accès sur le web (archives ouvertes, par ex. [MEDIHAL](#)) ?

UNE SEULE REPONSE !

- oui, au moins une fois
- oui, régulièrement
- non, je ne savais pas que c'était possible
- non, je ne sais pas le faire
- non, je n'ai pas le temps
- non, il faut un soutien pour le faire
- non, et je m'y [refuse service](#) ou du gestionnaire de site

18. Seriez-vous d'accord pour diffuser vos données en libre accès ? (Si diffusables et protégées juridiquement)

- oui, tout à fait
- oui, après un délai
- plutôt oui
- plutôt non
- non
- je ne sais pas

19. Parmi ces raisons, lesquelles vous inciterait à rendre vos données accessibles en libre accès ?

- une obligation du financeur (Horizon2020, ANR)
- la reconnaissance de la communauté scientifique
- une meilleure visibilité de vos travaux
- l'occasion de nouveaux contacts, coopérations
- le souhait de voir se développer d'autres recherches à partir de vos données
- le souci de pouvoir valider vos résultats
- les valeurs du Libre Accès
- des retombées financières (crédits de recherche ...)

20. Quelles seraient les raisons qui vous empêcheraient ou vous freineraient pour rendre vos données de recherche accessibles en libre accès ?

- des freins juridiques liés à la nature des données (données à caractère personnel, données sous droit, données sensibles)
- des raisons économiques (concurrence)
- des raisons scientifiques
- la crainte du plagiat
- des raisons personnelles (mes données m'appartiennent)
- le manque de temps
- le manque de connaissances dans la description des jeux de données, les modalités de dépôt ...
- le manque de compétence

21. Pour diffuser vos données en libre accès, quel est (ou serait) votre support principal ?

- une archive ouverte (HAL)
- une plateforme spécialisée dans l'archivage et la diffusion de la recherche
- un réseau social académique (ResearchGate, Academia ...)
- votre blog ou site personnel
- le site du laboratoire
- un serveur dans le laboratoire
- un data paper
- Autre : ...

VOS BESOINS, VOS ATTENTES

22. Savez-vous ce qu'est un Plan de Gestion de Données (PGD) ou Data Management Plan (DMP) ?

- Oui
- Non

23. Savez-vous ce qu'est une Très grande infrastructure de recherche (TGIR) ?

- oui
- non

24. Dans quelle mesure êtes-vous d'accord, ou non :

- je suis familier avec les enjeux de l'open access (accès ouvert) - oui – non
- je suis prêt à partager mes données avec les autres – oui – non
- je souhaiterais pouvoir stocker mes données au-delà de la durée du projet de recherche -oui –non
- je souhaiterais que le laboratoire dispose d'un plan de gestion de données –oui –non
- je souhaiterais que le laboratoire dispose d'un ensemble de métadonnées prescrit pour verser mes données vers un entrepôt –oui –non
- je souhaiterais que le laboratoire utilise un système de nommage standard de fichiers –oui –non

25. De quelles offres de services aimeriez-vous bénéficier ?

- des conseils relatifs à la publication et la citation des données de la recherche
- des conseils d'ordre général pour la gestion des données de la recherche

- des conseils sur des questions techniques (par ex. métadonnées, normes, archivage à long terme)
- des conseils juridiques (par ex. les restrictions d'accès, le traitement des données sensibles, l'utilisation de licences)
- une offre de formations sur la question des données de la recherche
- une assistance technique (par ex. pour la publication des données dans une revue)
- une aide pour préparer un plan de gestion de données
- un espace d'archivage sécurisé et fiable pour mes données
- autre :

Seriez-vous intéressé par la présentation au CRENAU des plateformes Humanum-Loire et PROGEDO ?

- oui
- non

Seriez-vous intéressé par une présentation au CRENAU d'un plan de gestion ?

- oui
- non