



HAL
open science

En archéologie, d'une ressource à une autre ou de l'interopérabilité pour tous

Véronique Humbert, Nathalie Le Tellier-Becquart, Blandine Nouvel

► To cite this version:

Véronique Humbert, Nathalie Le Tellier-Becquart, Blandine Nouvel. En archéologie, d'une ressource à une autre ou de l'interopérabilité pour tous. Open Science Practices : étudier les pratiques numériques au sein des communautés scientifiques, 2021. halshs-03927898

HAL Id: halshs-03927898

<https://shs.hal.science/halshs-03927898>

Submitted on 6 Jan 2023

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 4.0 International License

En archéologie, d'une ressource à l'autre ou de l'interopérabilité pour tous

Véronique Humbert (1), Nathalie Le Tellier-Becquart (2), Blandine Nouvel (3)

(1) Université Paul-Valéry Montpellier 3, CNRS, Ministère de la Culture, Archéologie des Sociétés Méditerranéennes, Montpellier, France

(2) Maison des sciences de l'homme Mondes, CNRS, Université Paris-Nanterre, Université Paris-Panthéon-Sorbonne, Nanterre, France

(3) Aix Marseille Université, CNRS, Centre Camille Jullian, Aix-en-Provence, France

Communication au colloque en ligne « Etudier les pratiques numériques au sein des communautés scientifiques » - Regard sur les outils et les données au prisme de la science ouverte, le 24 juin 2021.

La question de l'interopérabilité des données est essentielle pour l'archéologie. Parce que la discipline détruit par la fouille les preuves de ses analyses, la collecte d'information sur le terrain produit nécessairement une masse importante de données. Cahiers de fouilles, relevés, photos et photogrammétrie, analyses statistiques, dessins, etc. constituent une typologie riche qu'il s'agit d'organiser et de rendre disponible pour la communauté. De plus, la reprise fréquente de la documentation ancienne est aujourd'hui un enjeu qui rend l'urgence de la rationalisation cruciale. Le mouvement de la science ouverte vient conforter un processus engagé mais où le chercheur, au niveau individuel, se sent encore mal à l'aise, sinon peu concerné par des injonctions qui bousculent sa pratique.

Des expériences fédératrices

Pourtant, des expériences sont conduites par les institutions de recherche pour favoriser l'accès et le partage de ces données. Par exemple en archéologie, on citera les actions de Frantiq et celles du consortium MASA.

La Fédération et Ressources sur l'Antiquité ([Frantiq](https://www.frantiq.fr/), Groupement de services du CNRS) rassemble depuis 1985 plus de 40 structures, unités de recherche du CNRS et des universités, musées, services d'archéologie du ministère de la culture et collectivités territoriales. Sa politique documentaire, depuis ses débuts, met en œuvre l'accès libre, la conduisant à développer un [catalogue collectif](#) partagé pour les bibliothèques¹ avec un service de fourniture de documents, et un thésaurus spécialisé [PACTOLS](#), géré avec le logiciel de gestion de vocabulaire contrôlé [Opentheso](#)². Conformément à l'évolution du domaine, les outils de Frantiq sont normalisés et ouverts. Le Système de gestion de bibliothèque Koha et Opentheso fonctionnent en réseau. Pour favoriser l'accès aux

1 <https://www.frantiq.fr/>

2 <https://opentheso.hypotheses.org/>

ressources, ils associent, autour du thésaurus PACTOLS et du catalogue, des référentiels alignés (Wikidata, IdRef, Geonames, Insee, Art & Architecture thesaurus du Getty, Backbone Thesaurus de DARIAH-EU) et des réservoirs connectés (corpus de recherche EpiCherchell et SculptuRo pour commencer, bibliothèques numériques DIGIMOM et SRA Bretagne, revues de Mir@bel). L'interopérabilité repose sur des identifiants pérennes de type Ark attribués aux notices du catalogue comme aux concepts du thésaurus, produisant des métadonnées qui visent à respecter les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable).

Le [consortium MASA](#), Mémoire des archéologues et des sites archéologiques, créé en 2012 par la TGIR Huma-Num, œuvre à la diffusion des bonnes pratiques de la science ouverte dans la communauté des archéologues. Il a mis en place un écosystème particulier au cycle de vie de la donnée en archéologie qui articule des outils déjà en place comme Opentheso de Frantiq, des référentiels terminologiques (le thésaurus PACTOLS de Frantiq, Geonames, PerioDo), l'ontologie CIDOC-CRM, et en a développé de nouveaux comme la plateforme OpenArchaeo, OpenTermAlign, LogicistWriter. Tous sont interopérables et visent au partage et à la réutilisation des données en s'appuyant sur les normes et standards du web des données et sur l'alignement des données.

Ces initiatives sont l'œuvre de professionnels de l'information, d'archéologues ou d'informaticiens très impliqués dans la diffusion et le partage des connaissances, qui manipulent langages et concepts de l'informatique des données. Mais qu'en est-il sur le terrain, comment l'archéologue intègre-t-il dans sa pratique de recherche la dimension Science ouverte, comment saisit-il tous les enjeux des outils auxquels il est confronté ?

L'enquête

Le questionnaire adressé aux archéologues et aux historiens des périodes anciennes au premier trimestre 2021 par Frantiq avait pour objectif d'interroger leurs connaissances et leur utilisation des outils et des ressources documentaires dans le contexte désormais institutionnalisé de la science ouverte. Les réponses devaient permettre d'évaluer l'adéquation des outils et services de Frantiq. L'échelle d'observation de l'enquête se situe au niveau de divers établissements, dont le CNRS, l'INRAP, les universités, le ministère de la culture, les collectivités territoriales et les services d'archéologie publics ou privés.

Le temps de conceptualisation de l'enquête a été assez long. Il s'agissait de trouver le bon niveau de langage pour élaborer les questions, afin qu'elles soient compréhensibles par tous, en évitant d'utiliser le jargon métier des professionnels de l'information scientifique et technique. L'enquête a bénéficié de l'aide de Sylvie Grésillaud (INIST-CNRS) pour la mise en œuvre du questionnaire en ligne sur Webquest et de l'appui de divers réseaux pour assurer une diffusion qui soit la plus large possible auprès d'une communauté scientifique nationale estimée à environ 5000 archéologues issus de multiples institutions.

Nous avons identifié quatre thèmes répondant aux objectifs de l'enquête :

- Les pratiques documentaires (temporalité, collaborations, outils)
- Les ressources (bibliographie, référentiels, corpus et sources de données)
- La science ouverte (gestion des données, entrepôts d'archivage)
- Le soutien (formation, assistance, adéquation des outils et services).

Pour les besoins de cet exposé, et parce que le dépouillement du questionnaire n'était pas encore terminé lors de notre participation au colloque Open Science Practices de juin 2021, nous avons limité notre présentation à la partie des réponses portant sur la Science ouverte.

Les répondants

L'archéologue type : un fonctionnaire d'état titulaire d'un doctorat

393 personnes ont répondu au questionnaire. Les répondants sont pour 48 % fonctionnaires d'État, pour 10 % fonctionnaires dans la fonction publique territoriale. Ils sont rattachés quasiment à parts égales au CNRS (23 %), à l'INRAP (24 %), aux universités (21 %). 9 % travaillent au ministère de la culture. 26 % ont un autre statut et sont pour l'essentiel contractuels à l'Institut national de recherche archéologique préventive. 8 % sont des étudiants. 8 % sont salariés des opérateurs privés en archéologie préventive. Une majorité est titulaire d'un doctorat, 38 % ont un niveau bac + 5, mettant en évidence le haut niveau de qualification de la profession.

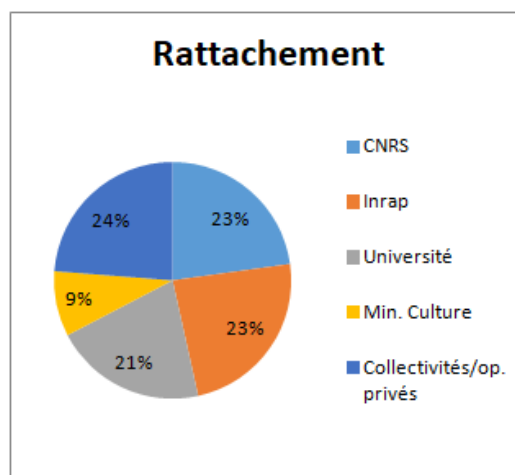


Figure 1 : Quel est votre organisme de rattachement ?

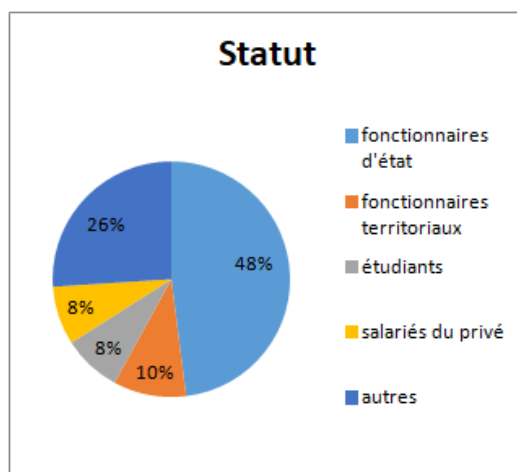


Figure 2 : Quel est votre statut ?

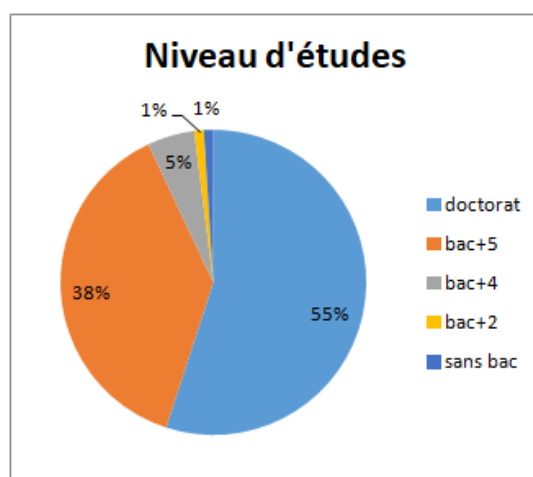


Figure 3 : Quel est votre niveau d'études ?

Le répondant-type est donc fonctionnaire de l'enseignement supérieur et de la recherche, personnel CNRS, universitaire ou INRAP et titulaire d'un doctorat.

Présence et identité numérique

Une prédominance des réseaux sociaux académiques privés

73% disposent d'une page personnelle sur internet via des sites ou blogs scientifiques de laboratoire, sur HAL ou des réseaux sociaux académiques sans qu'il soit possible de distinguer la part de chacun de ces canaux.

Les réseaux sociaux les plus utilisés par les sondés sont les réseaux sociaux de droit privé³ Academia.edu (dans 78 % des cas) et ResearchGate (5,21 % des réponses). 10 % des répondants ne sont présents sur aucun réseau social. On note que plus les répondants sont diplômés, moins ils vont sur les réseaux.

³ « Academia.edu est géré par l'entreprise américaine Academia (régie par le droit américain) et ResearchGate une société allemande (régie par le droit européen) »
<https://mae.hypotheses.org/author/letellier>



Figure 4 : Nombre de réseaux sociaux utilisés

IdHAL et ORCID, les identifiants numériques les plus connus

Si 61 % ont déjà créé leurs propres identifiants IdHAL ou ORCID⁴, 81 % ne connaissent pas leurs autres identifiants auteurs, tels qu'IdRef⁵, Vial⁶ ou ISNI⁷. Et de façon générale, 74 % ne savent pas à quoi servent ces identifiants. Parmi les 26 % qui ont affirmé en connaître l'utilité, seulement 14 % l'ont explicité. La création d'identifiants HAL est plus ou moins imposée par les institutions de recherche engagées dans la politique de science ouverte et ORCID de plus en plus réclamé par les éditeurs. Dans sa feuille de route publiée en 2019, le CNRS déclare que « seules les publications déposées dans HAL seront listées dans les rapports annuels d'activité des chercheurs du CNRS » avec l'objectif de 100 % de publications dans HAL d'ici fin 2020⁸. De nombreux éditeurs tels qu'Elsevier ou Nature⁹, ainsi qu'OpenEdition Center ont retenu l'identifiant ORCID qui devient ainsi incontournable pour tout publiant¹⁰.

La plupart des répondants ont compris que ces identifiants permettent la distinction des homonymes et la mise en relation de l'auteur avec ses publications. Peu conçoivent que ces identifiants permettent de lier leurs publications et leurs données. Les référentiels qui ne sont pas directement ancrés dans les pratiques des chercheurs, et notamment ceux utilisés par les bibliothèques (IdRef et Vial), sont peu connus. Or, plusieurs systèmes d'identification de personnes coexistent et pour en assurer l'interopérabilité, les auteurs doivent déclarer l'ensemble de leurs identifiants individuels sur chacun de leurs comptes.

Ces résultats, mis en regard avec ceux issus de l'étude réalisée en 2020 sur l'existence d'un profil numérique (ORCID, Researcher ID, Academia et ResearchGate) des 1047 chercheurs de toutes disciplines à l'université de Caen¹¹, sont néanmoins encourageants, puisqu'alors, moins de 40% des chercheurs de cette université avaient créé leur ORCID.

4 L'«Open researcher and Contributor ID» est un identifiant imposé par de nombreux éditeurs.

5 Produit de l'Abes, Agence bibliographique pour l'enseignement supérieur.

6 Virtual International Authority File, combine plusieurs fichiers d'autorités internationales produits par les bibliothèques nationales.

7 L'«International Standard Name Identifier» est une norme ISO utilisée par des bibliothèques, éditeurs, bases de données ou par les organismes de gestion des droits (artistiques notamment).

8 Entretien avec Alain Schuhl, directeur général délégué à la science du CNRS : <https://www.cnrs.fr/fr/cnrsinfo/notre-objectif-100-de-publications-en-libre-acces>, 21 novembre 2019.

9 <https://blog.impactstory.org/ten-things-you-need-to-know-about-orcid-right-now/>

10 <https://www.openedition.org/orcid> et <https://www.openedition.org/17989>

11 Boudry C, Durand-Barthez M (2020) Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia.edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study. *PLoS ONE* 15(9). <https://doi.org/10.1371/journal.pone.0238583>

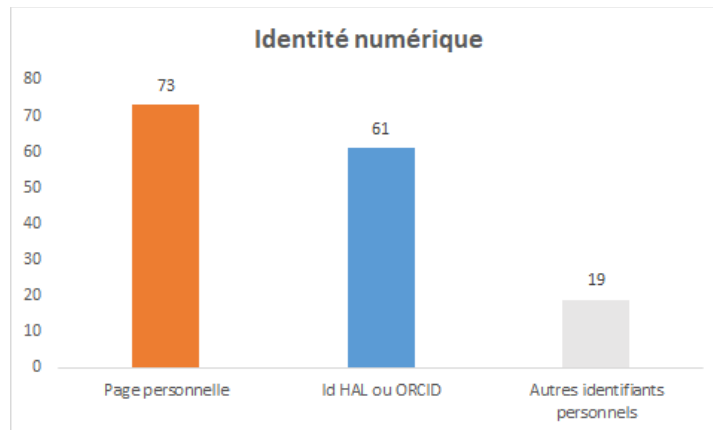


Figure 5 : Identité numérique des archéologues

Une minorité consciente de l'impact des moteurs de recherche

Les moteurs de recherche les plus utilisés par les répondants se répartissent ainsi : 86 % se servent de Google et 40 % de Google Scholar. 22 % recourent à Qwant et 13 % à Isidore. 16% des répondants se tournent vers d'autres moteurs de recherche comme Lilo¹², Ecosia¹³ ou Duck Duck Go¹⁴. Cette minorité, consciente de l'impact que peut avoir l'utilisation d'un moteur de recherche, pourrait contribuer à la sensibilisation des archéologues à ces problématiques de captation des données de navigation et des informations privées.

Ressources numériques : le plein texte plébiscité

83 % des sondés font bien la différence entre un catalogue de bibliothèque, une archive ouverte ou une plateforme de diffusion de publications scientifiques. Les ressources en ligne inscrites dans une politique d'ouverture des données, notamment géographiques, sont plébiscitées. Ainsi 78 % des répondants déclarent utiliser Géoportail, InfoTerre, Remonter le temps, cadastre.gouv.fr. Sans surprise, les ressources documentaires les plus consultées sont celles qui donnent accès au texte intégral des publications (Persée, HAL) avant les catalogues de bibliothèques spécialisées.

12 <https://www.lilo.org>

13 <https://info.ecosia.org/?tt=7e24f004>

14 <https://duckduckgo.com>

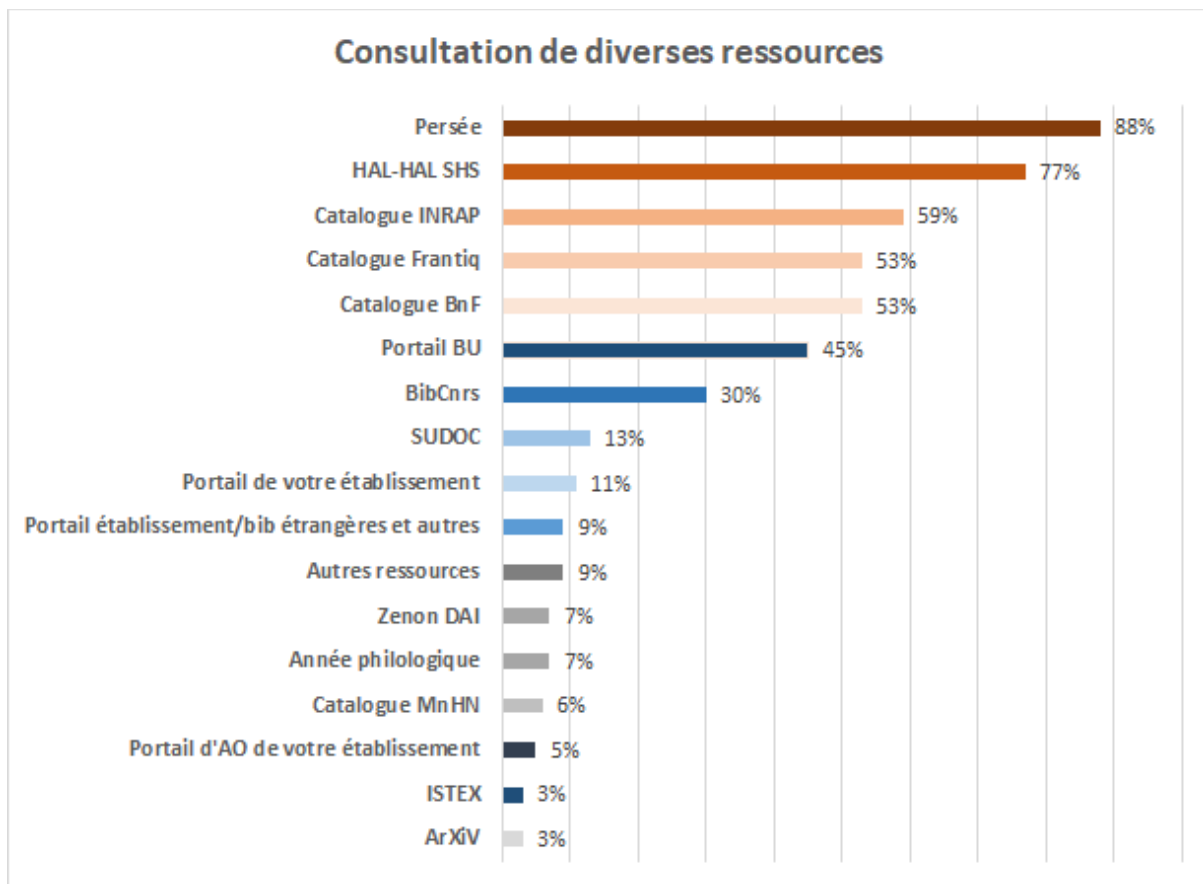


Figure 6 : Quelles ressources utilisez-vous ?

Pratique du dépôt/archivage : le dépôt des publications dans HAL majoritaire

71 % des répondants déposent dans une archive ouverte et le font à 85 % dans HAL. Malgré tout, il reste une confusion entre entrepôt et réseau social : parmi les archéologues qui déposent ailleurs que dans HAL, 48 personnes citent Academia¹⁵ et 23 ResearchGate. 57 % n'associent jamais de données à leur dépôt. 33 % ne savent pas que c'est possible, 20 % ne le font pas pour des raisons de propriété intellectuelle et 16 % de confidentialité. Les autres raisons invoquées par 25 % des répondants n'associant pas de données aux publications, sont le manque de temps, d'outils, d'intérêt ou d'utilité, ou encore l'absence de distinction entre données et publications.

¹⁵ Un répondant précise que HAL n'est « pas présent/connu dans les pays étrangers »

Les plans de gestion de données encore peu utilisés

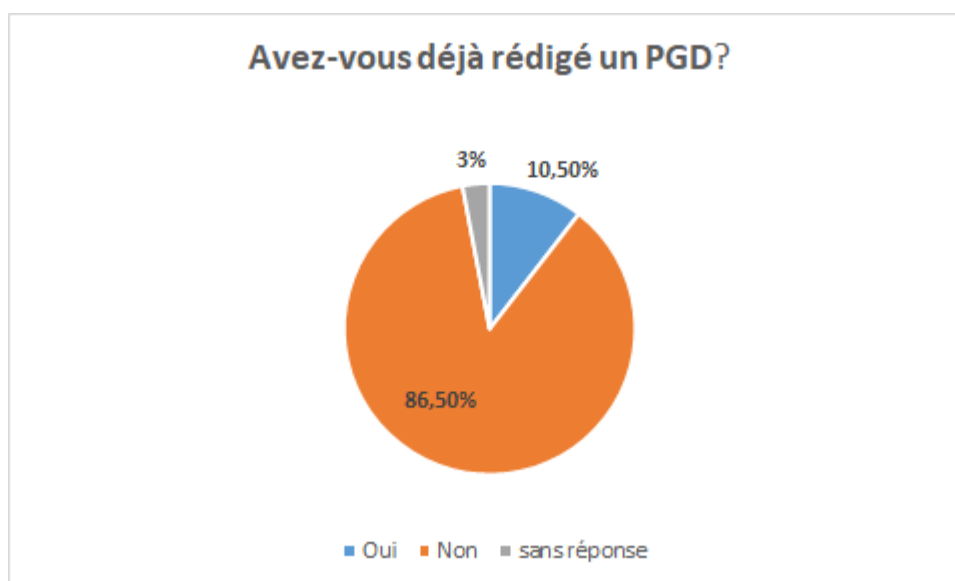


Figure 7 : Avez-vous déjà rédigé un plan de gestion de données (PGD) ?

84 % des répondants n'ont jamais rédigé de plan de gestion de données (PGD). Une large majorité ne sait pas ce que c'est. D'autres raisons sont invoquées, telles que le manque de temps, le fait de ne pas y avoir pensé, ou parce que « c'est trop compliqué ».

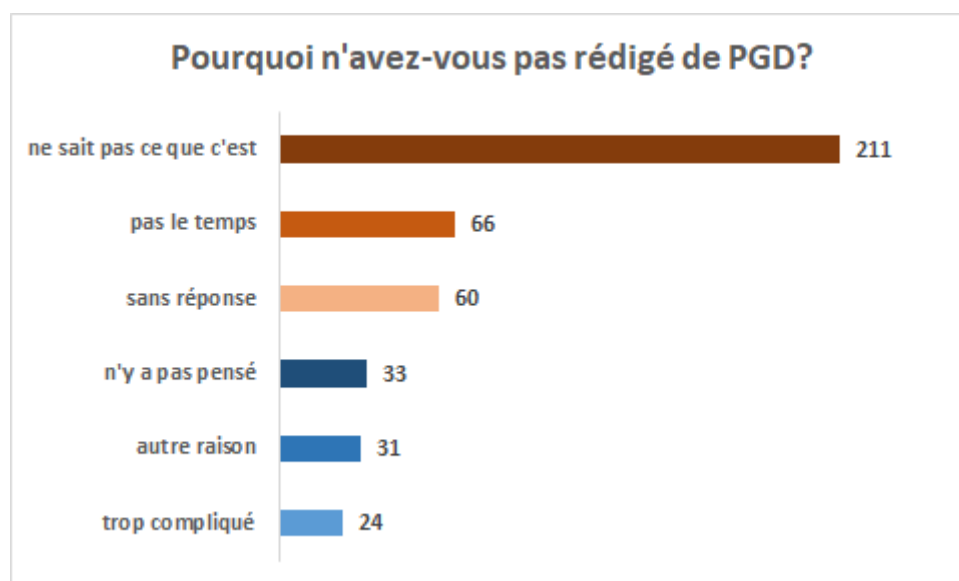


Figure 8 : Pourquoi n'avez-vous pas rédigé de PGD ?

Certains précisent qu'ils n'en ont pas besoin ou pensent que ce n'est pas de leur ressort¹⁶ tandis que d'autres y songent davantage¹⁷.

16 Par exemple : «C'est le porteur du projet qui a rédigé le plan de gestion des données, ce n'est pas dans mes attributions.»

17 Par exemple : «Je n'en ai pas eu le temps ni le besoin jusqu'à présent, j'aimerais m'y consacrer davantage...»

La partie très minoritaire des répondants ayant déjà rédigé un plan de gestion de données l'a fait soit en collaboration avec son équipe de recherche, ou seul, plus rarement avec l'aide d'un documentaliste ou d'un autre professionnel de l'information scientifique et technique (les archivistes sont mentionnés).

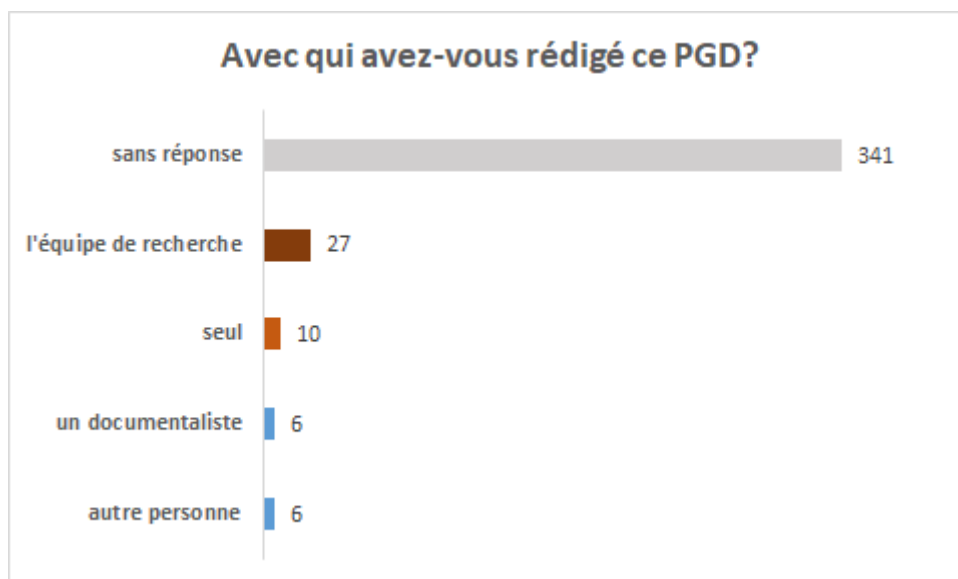


Figure 9 : Avec qui avez-vous rédigé un PGD?

Peu de modèles de plans de gestion de données sont cités. Le plus mentionné (4 personnes) est celui de l'Agence nationale de la Recherche. L'ANR demande à ce qu'un plan de gestion de données lui soit transmis dans les six mois après le démarrage des projets financés depuis 2019¹⁸. Le modèle du consortium MASA, Mémoires des Archéologues et des Sites Archéologiques¹⁹ n'est mentionné qu'une fois. Un répondant signale un "modèle fourni par le documentaliste" sans savoir lequel il s'agissait. On note dans les réponses une confusion entre modèles et outils d'aide à la création en ligne de PGD : les plateformes web DMPOpidoR²⁰ et DMPOnline²¹ sont citées comme modèles. Ce volet nécessite donc un travail de communication et d'accompagnement important.

18 <https://anr.fr/fr/actualites-de-lanr/details/news/lanr-met-en-place-un-plan-de-gestion-des-donnees-pour-les-projets-finances-des-2019/>

19 https://dmp.opidor.fr/template_export/942300517.docx

20 <https://opidor.fr/planifier/>

21 <https://dmponline.dcc.ac.uk/>

Accompagnement, formation

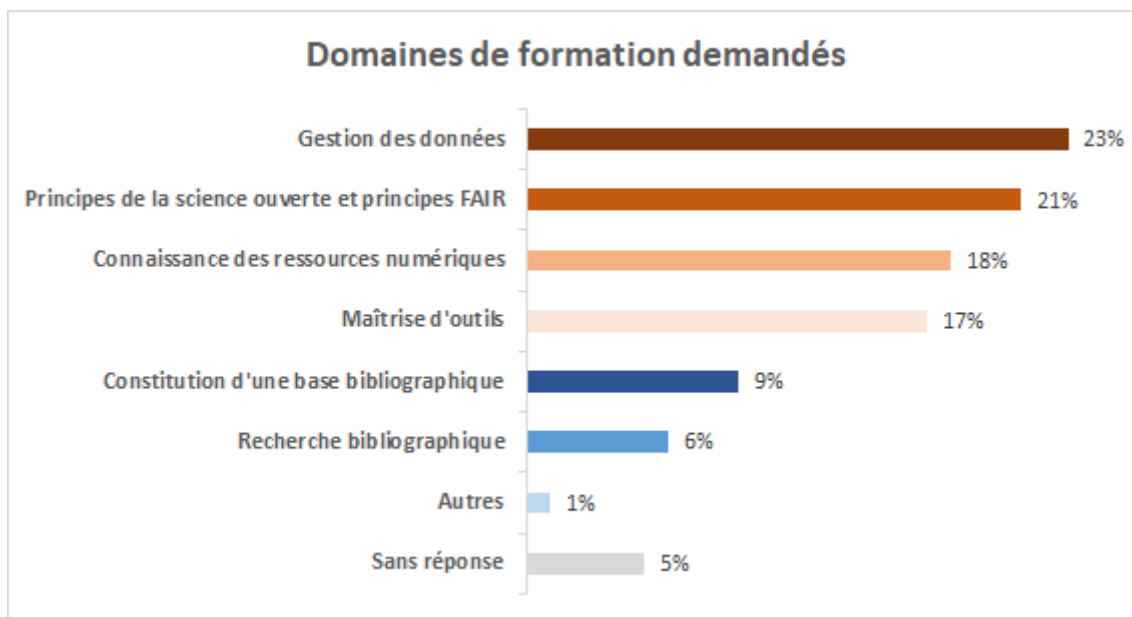


Figure 10 : Sur quel(s) domaine(s) souhaitez-vous des sessions de formation ou d'information ?

60 % des répondants demandent un accompagnement ou des formations autour des données, que ce soit pour la consultation de ressources documentaires ou archivistiques ou pour la gestion et le partage des données dans le cadre de la Science ouverte.

L'accompagnement souhaité est plutôt d'ordre technique que scientifique (13 %), il porte sur le partage de données et les nouveaux outils (13 %), et 7 % seulement concerne le dépôt dans les archives ouvertes, sans doute en résonance aux sensibilisations et formations qui ont lieu ces dernières années sur le sujet.

Un besoin subsiste toujours sur la formation à la recherche de ressources bibliographiques et d'archives.

Les autres requêtes relèvent plutôt de compétences en gestion bibliographique ou documentaire (thésaurus, indexation, références bibliographiques), de compétences informatiques (dont les bases de données) mais aussi sur la numérisation et le web sémantique, sans plus de précision. Quelque que soit leur demande, les répondants qui le précisent (4 %), signalent un manque de temps pour la réalisation de ces tâches, considérées souvent comme secondaires.

En perspective

Ces premiers dépouillements de l'enquête ont donc mis en exergue plusieurs points.

Le partage des données concerne plutôt les publications mais se conduit dans un contexte de forte incitation, sans réelle adhésion aux objectifs du mouvement de la Science ouverte. L'ouverture des données reste encore très modérée. De ce point de vue, les outils développés, tant par Frantiq que par le consortium MASA, et reposant sur les principes FAIR, trouveront un écho favorable auprès de la communauté lorsque la prise de conscience générale sera matérialisée par des dépôts et des mises à disposition de données par les producteurs eux-mêmes. Espérons de plus que l'effet de masse soit incitatif

tant l'intérêt d'utiliser des données déjà collectées est important. Il relève donc à ces organisations de maintenir leur effort de développement d'instruments et de services interopérables, complémentaires aux offres des infrastructures de recherche.

En particulier, Frantiq doit consolider l'accès vers le texte intégral depuis son catalogue en s'appuyant sur l'alignement de ses ressources, en particulier avec Mir@bel et IdRef pour les revues et les autorités. Sans doute faudra-t-il envisager un enrichissement automatisé avec HAL et Persée. Une expérience ancienne de rapprochement entre le CCI et Persée avait montré que le rapprochement automatisé des métadonnées restait délicat et peu opérationnel. Mais il faut compter désormais avec l'évolution des outils et l'important travail d'harmonisation et de nettoyage réalisé de part et d'autre. PACTOLS, promu comme référentiel terminologique pivot pour la discipline, constitue potentiellement un point de rebond vers des corpus de données, bibliographiques et scientifiques. L'interaction avec les bases de recherche, expérimentée avec des programmes du Centre Camille Jullian, EpiCherchell et SculptuRo, et finalisée cette année, valide le modèle. Les alignements avec les grands réservoirs que sont le Art&Architecture Thesaurus, Geonames, IdRef ou Wikidata, engagés opportunément dans le cadre de programmes nationaux CollEx-Persée ou internationaux comme ARIADNE Plus ou DARIAH sont évidemment à poursuivre.

En parallèle, Frantiq et MASA s'engagent de concert pour favoriser la montée en compétence des producteurs de données dont les besoins en formation se sont exprimés dans l'enquête Frantiq. Les documentalistes et les bibliothécaires développent depuis longtemps des actions en ce sens²². Les membres de Frantiq n'échappent pas à la règle : les tous nouveaux c@fés Frantiq participent de ce mouvement. Décidés lors de l'assemblée générale de 2021, ils se tiennent sur un format souple de 1 à 2 heures le vendredi matin, une fois par mois. Y sont traités des sujets/outils émergents qui demandent un besoin d'information ou de formation et sont un espace de sensibilisation à la Science ouverte. De son côté, le consortium MASA dispose d'une nouvelle offre de formation²³ élaborée en partie avec son partenaire Frantiq. Elle touche les outils et conduit le chercheur à mieux comprendre et mettre en œuvre l'ouverture de ses données dans le respect des bonnes pratiques.

Cette enquête aura un impact non seulement sur les évolutions des outils mis à disposition des archéologues par le GDS Frantiq, dans la continuité d'une politique documentaire ouverte initiée dès les années 80, mais elle est également susceptible de faire évoluer sa politique de formation en lien avec les actions du consortium MASA. 93 % des répondants fréquentent les bibliothèques et centres de documentation de leur institution, et ce, au moins une fois par mois. Ainsi, l'enquête démontre que le réseau Frantiq, grâce à son maillage territorial et inter-institutionnel, constitue un point d'appui incontournable pour accompagner et faire évoluer les pratiques des archéologues français.

22 voir notamment les ateliers Dialogu'IST de Renatis, la plateforme Doranum de l'Inist

23 <https://masa.hypotheses.org/formations>