Comment valoriser mes données produites à l'EFA ?

- Les données versées dans Archimage
- Les données déposées dans un entrepôt externe
- Comment valoriser les données à l'EFA

Les données versées dans Archimage

À l'EFA, le versement des documents par le chercheur au service des archives, avec un bordereau, implique un transfert de responsabilités : c'est l'établissement qui s'engage à les conserver et documenter selon les normes en vigueur. Concrètement, après avoir traversé les étapes de description, les données sont versées avec leur visuel et métadonnées associées dans l'entrepôt institutionnel (et donc développé et maintenu en interne) Archimage pour leur sauvegarde, en ce qu'elles représentent les travaux fournis par l'École. Le chercheur garde les droits moraux et patrimoniaux sur ses documents jusqu'à une éventuelle cession non-exclusive de droits pour partage avec la communauté scientifique. En effet, la mise à disposition se fait par le biais du service des archives, en accord avec les règles de diffusion et de communication ainsi que les principes de la science ouverte.

La plateforme Archimage n'est pas uniquement un lieu de gestion des données à visée archivistique, mais peut aussi être un espace d'exposition des documents versés à la photothèque, à la planothèque et aux archives manuscrites. Peuvent être recueillis tous types de données, allant du texte simple à la base de données plus complexe en passant par des analyses photogrammétriques (voir à ce sujet les fiches pratiques par typologie de document). En ce sens, elle permet une valorisation du patrimoine de la recherche à l'EFA qui ne peut être maintenue que par l'implication des chercheurs et membres associés dans la pratique de l'*open data*, soit l'accès et le partage libre pour tout le monde des données de la recherche, sous une forme normalisée et lisibles par des machines. Cette pratique se fait sous une certaine licence, de type Creative Commons, afin d'être en mesure de les transformer, combiner et partager, toujours en citant ses sources.

Les données déposées dans un entrepôt externe

À l'inverse du versement au service des archives par un bordereau, c'est le cas de figure du dépôt par le chercheur dans un entrepôt externe qui est présenté ici. Il en existe de plusieurs types et de différentes envergures, en voici quelques exemples :

- Recherche Data Gouv est un entrepôt interdisciplinaire et gouvernemental, mis à disposition par le ministère de l'enseignement supérieur et de la recherche pour ses communautés. Il propose une documentation fournie et notamment un outil d'aide à la rédaction de *data paper* pour les données qu'il accueille.
- Nakala est un entrepôt de données spécialisé dans les sciences humaines et sociales, mis à disposition par la TGIR (pour *Très grande infrastructure de recherche*) HumaNum. Il propose différents outils d'exposition notamment par le biais de son plug-in *Nakala-Press* qui permet de créer un site comme une vitrine pour les jeux de données accueillis.
- Zenodo, enfin, est une infrastructure à l'échelle européenne et interdisciplinaire qui propose un stockage sécurisé et pérenne pour des jeux de données académiques.

Ces entrepôts se caractérisent par le fait que le dépôts est effectué par le chercheur lui-même sur la plateforme, sans accompagnement de la part du service des archives. Il reste donc entièrement responsable de leur mise à disposition, par le choix ou non de l'application d'une barrière de consultation ou d'un embargo. Les entrepôts externes, de leur côté, se limitent à offrir une solution d'hébergement pérenne et sécurisé aux données, avec l'application d'un identifiant stable de type DOI, Handle ou ARK.

Les identifiants pérennes et référencements/liens établis entre les différents corpus sert de base au travail d'analyse des données, pour une exposition par le biais d'une publication ou d'une actualité sur un site internet, d'une base de données publiée ou d'un *data paper*. Une fois les résultats publiés, il est recommandé d'établir un lien entre la publication et les jeux de données entreposés grâce à l'identifiant pérenne, afin d'assurer une meilleure reproductibilité et capacité de citation.

> Attention cependant, il faut signaler que le dépôt dans un entrepôt externe n'affranchit pas le chercheur de verser ses archives à l'EFA

Comment valoriser les données à l'EFA

À l’EFA, un certain nombre d’outils sont à disposition afin de permettre aux chercheurs de réutiliser leurs données dans des travaux de recherche. Certains sont proposés en interne, d’autres sont développés en lien avec des partenaires. Ces outils sont là pour améliorer la connaissance que la communauté scientifique et le public ont des travaux de l'EFA mais aussi afin que les chercheurs puissent se réapproprier les données versées ou déposées en les appelant depuis leur entrepôt de stockage.

- Pour documenter la vie d’un projet de recherche, OpenEdition propose l’ouverture de carnets Hypotheses pouvant recevoir des billets d’actualité, des réflexions en cours, etc. Il s’agit d’un blog que les chercheurs peuvent alimenter en nouvelles et ressources en lien avec leur activité.
- Par le biais d’outils numériques, il est possible de concevoir un site plus complet pouvant accueillir par exemple une bibliothèque numérique (grâce aux CMS WordPress, Drupal ou Omeka) : ces sites-vitrines sont l'occasion de diffuser et d'exposer les données créées lors des travaux de recherche et de partager les premiers résultats ou analyses préliminaires.
- Les publications (Voir le catalogue des éditions de l’EFA) qui accueillent des travaux de recherche dans ses collections et revues. Ces publications se font la plupart du temps en *open access* sur les plateformes d’OpenEdition et de Persée. On peut y retrouver des données également versées dans Archimage et mises en interopérabilité par le biais d'un lien.
- La rédaction de *data papers* est encouragée par le ministère de l’enseignement supérieur et de la recherche : voir à ce sujet la fiche pratique *Qu’est-ce qu’un data paper ?*
- Enfin, le signalement des données stockées dans Archimage se fait en interopérabilité avec de grandes plateformes nationales et européennes de référencement pour pouvoir trouver et exploiter les documents mis à disposition : on peut citer parmi elles OpenAIRE, Ariadne plus, Calames ou France Archives. L'idée est de permettre une plus grande visibilité aux collections de l'EFA et de permettre, par les moyens numériques, une interopérabilité et une ré-exploitation optimisées.

On peut se référer à ce sujet au schéma explicitant le cycle de vie de la donnée à l’EFA, mettant en lumière les étapes entre le dépôt d’une donnée et la mise en valeur des résultats acquis.

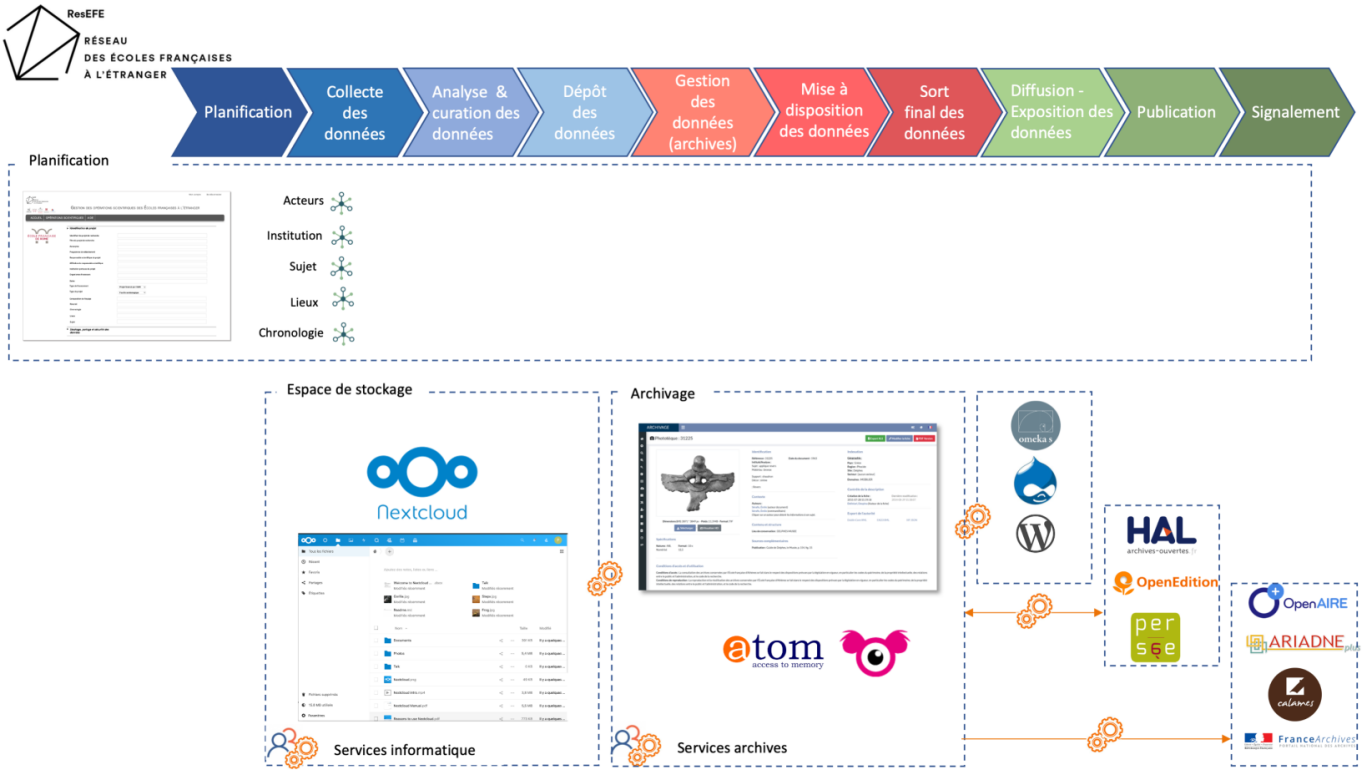


Figure : le cycle de vie de la donnée à l'EFA (crédit Bruno Morandière, ResEFE)

Ici sont exposés tous les outils et les acteurs permettant de mettre en avant la recherche effectuée ou en cours à l'EFA. On y voit bien que le versement ou dépôt sur un service d'entrepôt de données n'est pas une fin en soi, mais une étape vers la valorisation du travail des chercheurs en lien avec la communauté scientifique et les infrastructures de recherche partenaires. L'exposition des données est ce qui permet de préparer leur publication, et le signalement l'amorce d'une réutilisation potentielle dans le milieu académique ou de l'enseignement.

Ces possibilités de valorisation impliquent que les données aient été versées sous une licence libre, permettant leur partage dans les termes de l'*open data*. Il s'agit également d'être vigilant aux métadonnées associées à tout document, car cette documentation de la donnée est ce qui garantit sa qualité et son intégrité. Il convient donc d'y prêter attention dès le moment de la collecte des données, au début du projet (Voir à ce sujet la fiche pratique *Qu'est-ce qu'une donnée de qualité ?*).