

STRATÉGIE OPEN RESEARCH DATA

I. OBJECTIFS DE LA STRATÉGIE

Le mouvement Open Science vise à démocratiser les connaissances, via l'accès libre aux publications de la recherche scientifique (*Open Access*), au code source des logiciels (*Open Source/Open Code*), et aux données utilisées pour produire les connaissances (*Open Research Data, ORD*). Ouvrir les données de la recherche, c'est permettre leur utilisation par des personnes extérieures aux équipes de recherche afin de produire de nouvelles connaissances, d'améliorer la transparence de la science, et d'économiser l'argent des contribuables. Cela implique de gérer ces données (*data management*), les préparer, et les documenter dans ce but tout au long du projet de recherche.

En prenant le parti de ne pas limiter les données de recherche à une seule vie, mais de rendre de nouvelles utilisations possibles, l'ORD modifie la façon de penser les données au sein de la communauté académique et scientifique. Il est donc vital pour les structures de recherche de se doter d'une stratégie en matière d'ouverture des données, et d'offrir recommandations et soutien pour accompagner leur communauté à ces nouvelles pratiques. Ce document vise à clarifier la position et les actions de l'UniNE en matière d'ORD.

A. Proposer une vision cohérente

En dépit de sa taille, l'UniNE est dotée d'une riche diversité de disciplines et de thèmes de recherche. Cela signifie que les données qui y sont récoltées varient énormément en nature et en modalités d'acquisition, mais aussi que des contraintes particulières (droit d'auteur, embargo, données sensibles...) peuvent s'appliquer à certains projets. Si cette diversité est une force pour l'UniNE, elle implique des pratiques différentes à la fois concernant la gestion, la curation et la diffusion des données, et rend impossible toute standardisation des procédures.

Cela n'empêche pas l'UniNE de se donner une orientation générale, et de présenter une vision cohérente et commune face aux enjeux de l'ORD. Cette orientation stratégique permet d'officialiser les intentions de l'UniNE, et les moyens qu'elle met en place pour les accomplir. A l'échelle suisse, cela améliore sa visibilité, et participe à la transition du pays vers le mouvement mondial de l'Open Science.

B. Pouvoir recenser les ORD

Les données produites par les équipes de l'UniNE sont un bien commun financé par des fonds publics, dont la valorisation est bénéfique à la fois à la société, à la réputation de l'Université, ainsi qu'aux chercheurs et chercheuses. La stratégie ORD vise à rendre ces données trouvables, identifiables, et réutilisables. Si se doter d'une banque de données institutionnelle n'est pas possible en l'état, recenser les lieux

où les équipes diffusent leurs données permettra de rendre compte de la production scientifique de l'UniNE.

Recenser les données implique un travail à trois niveaux :

- Au niveau des jeux de données :
 - Attribution d'identifiants uniques et pérennes (DOI)
 - Renseignement des métadonnées sur le projet de recherche
- Au niveau des répertoires de données, ou entrepôts de données, ou *data repositories* (DR) :
 - Utilisation de data repositories [adaptés et reconnus](#)
 - Utilisation d'identifiants personnels pour les membres des équipes de recherche (ORCID)
 - Indiquer systématiquement l'UniNE dans les affiliations
- Au niveau de l'UniNE :
 - Disposer de recommandations au sein de chaque faculté/institut pour privilégier l'utilisation cohérente d'une ou plusieurs plateformes
 - Disposer de pages internet recensant les différents DR utilisés par les équipes de l'UniNE

C. Permettre la réutilisation des données

Publier les données poursuit deux finalités principales : leur préservation et leur diffusion à des fins de réutilisation. Cela permet de rendre la recherche scientifique transparente, en améliorant la reproductibilité, ainsi que plus économique, puisque les données peuvent être rentabilisées par de multiples utilisations.

La réutilisation n'est cependant complètement possible que si les données respectent les [critères FAIR](#) :

- *Findable* : les données doivent être trouvables, ce qui implique qu'elles soient publiées, identifiées (DOI) et référencées.
- *Accessible* : les données et métadonnées doivent être accessibles en ligne.
- *Interoperable* : les données et métadonnées doivent pouvoir être utilisées conjointement avec d'autres données, par le biais de formats, standards et vocabulaires contrôlés communs.
- *Reusable* : les données doivent être documentées et contextualisées de façon à pouvoir être réutilisées.

Pour que les données de recherche soient intelligibles et réutilisables, plusieurs actions doivent être mises en œuvre :

- Documentation des données : cela correspond à toutes les informations qui vont permettre de comprendre les données, depuis leur production jusqu'à leur publication. Cela permet de savoir quoi chercher et comment le chercher, et d'éviter de mettre à disposition des données inutilisables, ou pire, qui pourraient être utilisées de façon erronée.
- Curation des données : inutilisables en l'état, sensibles, difficilement documentables... Toutes les données n'ont pas vocation à être diffusées. Il appartient à l'équipe de recherche, avec le conseil de l'équipe ORD, de déterminer ce qu'il est nécessaire et possible de diffuser, ainsi que où et comment.
- Gestion des formats de données : si les logiciels libres sont à privilégier, chaque champ de recherche a des particularités et des outils propres à prendre en considération. La question des formats de préservation est aussi à prendre en compte.

- Publication des données : implique le choix d'un Data Repository adapté au besoin du projet de recherche, ainsi qu'un référencement efficace à l'aide de métadonnées.

D. Eviter les pièges de l'Open Data

Du moment qu'elle respecte certains principes, l'ORD est un outil puissant pour la recherche. En tant que telle, elle doit être utilisée au service des équipes.

Quelques écueils et difficultés auxquels la stratégie de l'UniNE se doit d'apporter des solutions :

- Dissémination désordonnée : pas de politique commune au sein des équipes en matière de dépôt, données publiées mais introuvables, ou publiées à plusieurs endroits...
- Cimetières à données : plus qu'un lieu, c'est une façon de déposer les données à la va-vite, sans documentation ni considération pour leur réutilisation, qui enlève tout son intérêt à l'open data.
- Processus rigides : vouloir à tout prix utiliser une façon de faire identique pour des besoins différents.
- Perfectionnisme : la documentation des données peut être chronophage. Les moyens mis en place pour la diffusion des données doivent être proportionnels avec le potentiel de réutilisation et tenir compte des possibilités pratiques des équipes de recherche, ainsi que des ressources informatiques nécessaires à la publication.

II. LA STRATÉGIE ORD

L'UniNE promeut une diffusion des données de recherche aussi ouverte et exhaustive que possible. Elle engage ses équipes de recherche à respecter la législation suisse et les exigences des bailleurs de fonds. Elle choisit SWISSUbase comme partenaire privilégié pour la diffusion des données, et conseille son utilisation, sans l'imposer à ses équipes. Elle affirme la diversité des approches et des disciplines comme une force, et met en place un service de data stewardship pour accompagner de façon personnalisée les équipes dans leurs démarches.

Cette stratégie repose sur trois piliers :

- La sensibilisation et la formation à l'ORD
- Le modèle de diffusion des données
- L'accompagnement personnalisé comme norme

A. Sensibilisation et formation

Bien qu'ayant connu un gain de popularité considérable ces dernières années, l'Open Science reste une approche nouvelle, qui implique un apprentissage. Du Data Management Plan aux entrepôts de données en passant par les données sensibles, de nombreux concepts et exigences peuvent désarçonner les équipes de recherche, ou rencontrer une certaine résistance de leur part en raison de la charge de travail supplémentaire impliquée. La sensibilisation passe par l'accès aux informations nécessaires (documentation en ligne) et par le contact direct (participation aux conseils d'instituts et de facultés, rencontres directes avec les membres de l'UniNE...)

Une [documentation en ligne complète](#) est hébergée sur le site des bibliothèques de l'UniNE. Son but est de présenter les enjeux de l'ORD de façon synthétique, car la plupart des chercheurs et chercheuses ne se confrontent à ces thématiques qu'au moment de publier des données, et ne cherchent pas un guide

complet mais des réponses rapides à leurs questions. Pour plus d'informations, cette documentation renvoie aux documents officiels (textes de loi, recommandations du SNF) et à des guides thématiques.

La documentation aborde entre autres :

- La présentation de l'Open Science et l'Open Research Data
- Le cycle de vie des données
- Les obligations légales concernant les données de recherche
- Les Data Management Plan
- La gestion des données au cours des projets (stockage, organisation, sécurité)
- La diffusion des données (plateformes, licences...)

Deux équipes participent activement à l'intégration de ces considérations dans les projets de recherche :

- L'équipe du Soutien Recherche et Innovation (contact.sri@unine.ch) apporte son soutien sur ces thématiques dès la genèse des projets de recherche, notamment lors des recherches de financement.
- L'équipe Open Research Data (contact.ord@unine.ch) accompagne les chercheurs et chercheuses dans le data management, la documentation et la diffusion des données de recherche.

L'apprentissage de l'Open Science se fera aussi au moyen de formations et séminaires, de façon à promouvoir cette pratique et accompagner à la transition. Des interventions régulières dans les laboratoires et instituts sont pratiquées, et la création de supports de cours spécifiques intégrant l'ORD est en projet.

B. Le modèle de diffusion des données

Le partenaire officiel pour la diffusion des données de l'UniNE est [SWISSUbase](#). En effet, cette solution présente des avantages indéniables :

- Partenaire suisse, indépendant d'éventuels changements de régulations à l'étranger
- Visibilité au sein de la recherche suisse
- Utilisation par d'autres universités (UNIL, UZH)
- Documentation extensive des données
- Métadonnées harmonisées en respect de standards internationaux
- Equipe de soutien (Data Service Unit)
- Moissonnage par OAI-PMH

Cependant, il est évident que cette solution ne conviendra pas à tous les projets, pour des raisons diverses :

- Utilisation d'un entrepôt disciplinaire sollicité par un champ de recherche particulier
- Engagement du projet à diffuser les données sur un autre entrepôt
- Habitude ou simple préférence individuelle pour un autre entrepôt
- Faible potentiel de réutilisation des données incompatible avec l'investissement de temps nécessaire pour publier les données sur SWISSUbase

L'UniNE propose donc un modèle à trois possibilités. In Fine, ce sont les équipes de recherches, et plus précisément le ou la *Principal Investigator* qui décident où seront diffusées les données. Aucune obligation ne leur est imposée de la part de l'UniNE. En revanche, les éditeurs scientifiques peuvent avoir des exigences concernant le DR où les données devront être publiées.

- SWISSUbase : la solution privilégiée par l'UniNE. Implique un accompagnement, ainsi qu'un certain investissement des équipes.
- Data Repositories disciplinaires : si la visibilité des données auprès des collègues d'un champ de recherche est mieux assurée par le recours à un DR reconnu par cette communauté, le recours à ce DR est recommandé.
- Data repositories généralistes : ils sont souvent la solution la plus facile pour les chercheurs et chercheuses. En conséquence, certains sont parfois utilisés pour déposer les données rapidement, sans considération pour le référencement et la description de l'étude et/ou des données. Si leur utilisation n'est pas déconseillée par l'UniNE, il convient de garder à l'esprit l'importance de possibles réutilisations des données lorsqu'ils sont choisis.

C. L'accompagnement personnalisé comme norme

Comme expliqué au point [A](#), la documentation mise à disposition de la communauté de l'UniNE a vocation à fournir une information concise, et non pas un panel exhaustif de tous les cas de figures qui peuvent se présenter lors de la diffusion de données à des fins de préservation et de mise à disposition. Chaque projet étant unique, il est vain d'espérer résoudre tous les problèmes qui se présenteront aux équipes de recherche par un labyrinthe de documentation.

En conséquence, les équipes ORD et SRI fournissent un accompagnement personnalisé sur demande aux chercheurs et chercheuses de l'UniNE, tout au long de leurs projets :

- Recherche de financement (SRI)
- Elaboration des projets de recherche (SRI)
- Droit et propriété intellectuelle (SRI)
- Data Management Plan (ORD)
- Organisation et sauvegarde des données (ORD)
- Curation de données (ORD)
- Plateformes de diffusion (ORD)
- Licences d'utilisation (ORD)

Cet accompagnement direct est un contrepoint nécessaire à la documentation en ligne, mais repose sur la volonté des équipes à y recourir.

III. CONCLUSION

Ce document décrit la stratégie de l'UniNE en matière d'ouverture des données de la recherche à un instant donné. Il forme le texte de référence à ce sujet. Il est voué à évoluer en accord avec les législations et recommandations officielles, ainsi que selon les retours et besoins de la communauté académique de l'UniNE.