

Projet « *HRS4R – Research Data Management* »

Faciliter la gestion ouverte et responsable des données de la recherche

Avec le soutien de
la Fédération
Wallonie-Bruxelles

Judith BIERNAUX
Isabelle HALLEUX
avec la collaboration de
Adeline GRARD

31 Mai 2022

Table des matières

Introduction : gestion ouverte et responsable des données en Fédération Wallonie-Bruxelles	3
Contexte local et européen	3
Le projet « HRS4R – Research Data Management ».....	5
Guidance : les outils et infrastructures	6
Actions accomplies	6
Prochaines étapes	7
Sensibilisation : conscientiser les utilisateurs des outils.....	7
Actions accomplies	7
Prochaines étapes	9
Formation : connaissances et autonomie	10
Actions accomplies	10
Prochaines étapes	11
Perspectives et conclusion	11
Guidance.....	11
Sensibilisation.....	11
Formation	11
Remerciements	12
Bibliographie	12
Annexe 1 - Résumé des livrables.....	13
Annexe 2 - Guide du Data Ambassador en FWB.....	14

Introduction : gestion ouverte et responsable des données en Fédération Wallonie-Bruxelles

Contexte local et européen

Les universités de la Fédération Wallonie-Bruxelles cultivent des valeurs de **qualité** et d'**excellence** en recherche, qui s'accompagnent d'une éthique et d'une intégrité aussi irréprochables que possible. En particulier, la **reproductibilité** scientifique est au cœur de leurs préoccupations.

La digitalisation des méthodes de recherche a déjà œuvré ces dernières décennies à normaliser l'ouverture des publications scientifiques, pratique que le **décret Open Access de la Fédération Wallonie-Bruxelles** de 2018 a définitivement ancrée, notamment via l'utilisation de dépôts institutionnels. D'autres objets scientifiques cheminent vers cette ouverture progressive, via les pratiques d'Open Software ou d'Open Edition, par exemple. La même année, le Groupe de Suivi de l'Espace Européen de la Recherche en FWB (GSEER) a mis en place un Groupe de Travail sur l'Open Science (GTOS), qui a proposé une **feuille de route pour l'Open Science en Fédération Wallonie-Bruxelles**. Cette feuille de route mentionne l'Open Research Data, les **principes FAIR** [1], et les **plans de gestion de données** comme actions nécessaires pour renforcer cette culture d'ouverture (cf. Tableau 1). Ces pratiques favorisent grandement le retour sur l'investissement public en facilitant la réutilisation de données, et permettent de répondre au moins en partie à la **crise de la reproductibilité scientifique** [2, 3, 4]. En effet, l'Open Data et les principes FAIR aident à contrer les mauvaises habitudes qui peuvent être imprimées dans les comportements de chercheurs mal sensibilisés subissant une importante pression de publication (*publish or perish*) [5, 6]. Parmi ces mauvaises habitudes, l'on retrouve par exemple le *p-hacking*¹, la publication sélective, le *data dredging*², voire la falsification de données : des pratiques qui se verraient menacées par une application plus généralisée de la mise à disposition systématique des données accompagnant chaque publication [7].

Tableau 1 : actions de la feuille de route pour l'Open Science en FWB pour lesquelles le présent projet permet de répondre.
En gras, les actions identifiées comme prioritaires par le Conseil des Recteurs.

12a	Rendre obligatoire la rédaction d'un DMP (Data Management Plan), selon des modèles uniformisés et rationalisés, notamment via l'outil dmponline.be, par les bénéficiaires de subventions du F.R.S.-FNRS, tout en encourageant l'ensemble des chercheurs financés par des fonds publics à en faire de même
12b	Réfléchir à rendre obligatoire, de façon plus générale, l'application des principes FAIR dans la gestion – et le cas échéant le partage – des données dans le cadre de recherches financées par des fonds publics
13	Définir, mettre en œuvre et communiquer des politiques institutionnelles explicites en matière de structuration, d'accessibilité et d'ouverture des données
14	Par défaut, assurer la publication des données au plus tard à la publication de l'article scientifique. Le cas échéant, le DMP mentionnera les raisons pour lesquelles il est dérogé à ce principe [...]
15	Intégrer les contraintes liées au RGPD dès la phase de production des données.
17	[...] Mettre en place et financer les infrastructures, les services et les outils nécessaires au déploiement de l'archivage, de la publication et du partage des données selon les normes FAIR en Fédération Wallonie-Bruxelles et en Wallonie

¹ Sélectionner artificiellement les combinaisons de tests et de variables qui donnent des résultats statistiquement significatifs, même s'ils sont aléatoires

² Altérer certains paramètres d'un échantillon de données pour obtenir artificiellement un résultat statistiquement significatif, même s'il est aléatoire

18	<i>Dans la mise en place de ces infrastructures, services et outils, veiller à</i> • <i>éviter toute duplication avec des infrastructures non-commerciales déjà existantes et répondant aux normes FAIR, utilisées ou utilisables par les chercheurs de la Fédération Wallonie-Bruxelles ;</i> • <i>assurer le meilleur équilibre et la meilleure complémentarité possibles entre</i> ○ <i>les infrastructures déjà existantes et utilisées par les chercheurs et les infrastructures nouvelles à mettre en place ;</i> ○ <i>les infrastructures centralisées aux échelons régional ou national (par ex. SODAH) et celles développées et entretenues au niveau institutionnel (universités et/ou agences de financement) ;</i> ○ <i>services et outils mutualisés (par ex. : anonymisation des données, formations et tutoriels, ...) et spécifiques ;</i>
21	<i>Encourager les chercheurs à exploiter, à tous les stades de leur recherche, des solutions Open Science non propriétaires et non commerciales afin de ne pas être captifs d'un fournisseur unique, ou d'un oligopole de prestataires de service ;</i>
22	<i>Stimuler la collaboration, en matière d'infrastructures et de services, entre la Wallonie, la Fédération Wallonie-Bruxelles et les autres entités fédérale et fédérées</i>
23	<i>Développer et entretenir activement une culture Open Science dans les institutions d'enseignement supérieur de la Fédération Wallonie-Bruxelles, en collaboration avec toutes les parties prenantes impliquées et en veillant à faire connaître les outils existants déjà disponibles.</i>
29	<i>Assurer un accompagnement des chercheurs dans la préparation du plan de gestion des données et informer ces derniers sur la possibilité d'inclure les frais liés à la gestion et au partage des données dans les budgets des projets (en fonction du bailleur de fonds)</i>

En parallèle, de nombreux acteurs scientifiques européens accordent de plus en plus d'importance à l'ouverture des données de la recherche. L'on peut citer l'exigence dans le **nouveau programme européen Horizon Europe** de fournir des plans de gestion de données comme livrables dans les projets de recherche. La plateforme [European Open Science Cloud](#) (EOSC) témoigne également d'une volonté européenne d'ouverture de la science.

Il appartient aux institutions de sensibiliser les chercheurs à ces problématiques et de les outiller au mieux pour leur permettre de répondre à ces nouvelles exigences. Ces outils doivent cependant être adoptés au-delà de la simple conformité et permettre d'instaurer les principes FAIR comme **habitude de travail** dans la communauté scientifique de Fédération-Wallonie Bruxelles. Si depuis le décret Open Access, la publication ouverte des articles scientifiques est bien imprimée dans le travail des chercheurs en Fédération Wallonie-Bruxelles, il n'en va pas de même pour le partage systématique et efficace (FAIR) des données associées à ces publications, et une ouverture à deux vitesses s'installe entre les résultats et les données qui permettent de les obtenir et de les vérifier. En revanche, ce contraste est moins présent dans d'autres pays européens, comme par exemple aux Pays-Bas, où de nombreuses infrastructures nationales³ favorisent le FAIR data.

Stratégie HR Excellence in Research (HRS4R)

Par sa dynamique de **concertation** pour l'amélioration de leurs conditions de travail, la méthode de travail déployée dans le cadre de la politique [HRS4R](#) a permis de structurer le travail présenté ici dans un programme-cadre. Cette méthode, éprouvée par des audits réguliers, repose sur une collaboration entre services institutionnels (notamment administrations) et **chercheurs**, afin de garantir la pertinence des actions prises auprès de leurs bénéficiaires directs. L'amélioration de l'environnement

³ On peut citer par exemple le répertoire de partage de données [4TU.ResearchData](#), ou encore la coalition d'archives de données de recherche [Research Data Netherlands](#) (RDNL), ainsi que le réseau [Data Archiving and Networked Services](#) (DANS).

de travail, notamment via les **outils**, **services** et **formations**, étant un pilier de la stratégie HRS4R, le projet dont il est ici question s'inscrit parfaitement dans son contexte.

Parmi les éléments de cette méthode, l'on retrouve par exemple l'idée d'une *gap analysis*, permettant de visualiser le pont à construire entre la situation actuelle, ou plutôt ses manquements, et la situation idéale exprimée par les chercheurs. Plus précisément, il s'agit de traduire les besoins de ces derniers en actions sur base d'un état des lieux en un temps donné. Dans le cadre de ce travail, les *gaps* identifiés ont permis de dégager les besoins généraux suivants :

- Recommandations pour des politiques institutionnelles de **gestion** responsable, éthique et ouverte des données de la recherche
- Centralisation des informations et ressources de gestion des données à disposition du chercheur en un point de contact aussi unique que possible, en formalisant le rôle de **Research Data Officier** (RDO)
- Mise en réseau des RDOs, centralisation et diffusion de supports de guidance cohérents entre institutions (modèles, procédés et recommandations) pour la rédaction de **plans de gestion de données** et le dépôt de données en ligne
- Recommandations pour la **formation** des chercheurs et des **personnes-ressources** au sujet de la collecte, l'utilisation, le stockage, la préservation et le partage des données de la recherche conformément aux principes FAIR, en tenant compte des spécificités disciplinaires

Ces grandes lignes ont permis aux bénéficiaires de ce projet de structurer les objectifs du projet selon trois axes de travail : Guidance, Sensibilisation, Formation.

Le projet « HRS4R – Research Data Management »

Le projet *HRS4R – Research Data Management*, noté ci-après « projet RDM », a été entériné par l'Arrêté du Gouvernement de la Communauté française du 10 décembre 2020. Il a impliqué toutes les universités de la Fédération Wallonie-Bruxelles, sous la coordination de l'ULiège. Il était initialement prévu du 1^{er} octobre 2020 au 31 décembre 2021, mais à la demande de l'ULiège et suite aux ralentissements liés à sa mise en œuvre, une prolongation a été accordée jusqu'au 28 février 2022. Son objectif était de **renforcer le soutien aux chercheurs en gestion responsable des données de la recherche**.

Une équipe de sept personnes représentant les six universités francophones belges s'est emparée de l'exécution de ce projet. Leur première action a été d'envisager le travail sous l'axe des **livrables**⁴, en dégageant trois axes de travail : **la guidance, la sensibilisation et la formation**.

⁴ Les livrables précis et leur état d'avancement sont repris dans l'annexe 1.

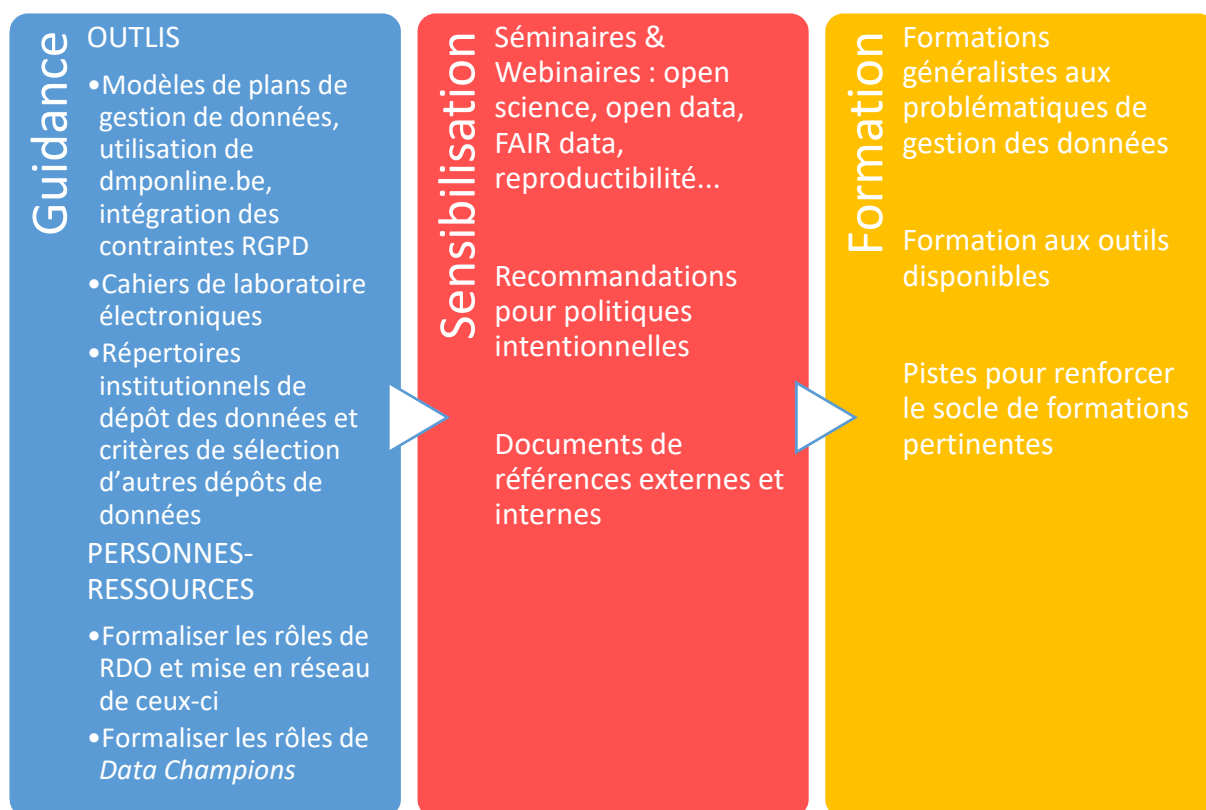


Figure 1 : Résumé des trois axes de travail du projet

Guidance : les outils et infrastructures

Actions accomplies

La base de travail choisie pour ce projet était la formalisation du rôle de Research Data Officer, adopté par les membres de l'équipe-projet, qui se sont renseignés comme tel auprès des chercheurs de leurs institutions. Par des réunions mensuelles, un **réseau interuniversitaire de RDOs** a pris forme, au sein duquel la circulation des informations et des idées a permis à chaque institution de profiter de l'inspiration et de l'avancement des autres pour améliorer le niveau de maturité de ses outils et services en gestion des données de la recherche.

La mise en réseau des RDOs a également permis de **rassembler** les différentes initiatives institutionnelles pour une meilleure **coordination** : par exemple, le travail réalisé dans le cadre du projet MQTOS a pu être suivi de près et a conduit à une accélération, pour les institutions qui accusaient un retard, du développement d'un dépôt institutionnel de données. À l'heure actuelle, ces développements, bien que toujours en cours, bénéficient de l'expertise partagée des intervenants (informaticiens, bibliothèques, ...) des différentes institutions, grâce aux discussions catalysées par le réseau de RDOs. Plus indirectement et grâce à une démarche concertée avec les *Data Protection Officers* (DPOs), l'intégration des contraintes [RGPD](#) dans l'outil de rédaction de plans de gestion de données [DMPonline.be](https://dmponline.be) (action 15 de la feuille de route pour l'Open Science en FWB) a pu aussi bénéficier de l'appui du réseau de RDOs et est aujourd'hui complète ou quasiment complète dans toutes les institutions FWB. Enfin, la participation régulière de certains RDOs au GT Open Science du Conseil des Recteurs permet un **suivi** régulier par les autorités interuniversitaires des actions proposées.

En somme, **le paysage universitaire francophone de la gestion des données de la recherche a trouvé un cadre structurant et efficace** qui permet d'avancer de manière cohérente.

Parmi les outils spécifiques sous la responsabilité (partagée ou non) des RDOs, l'on retrouve :

- [DMPonline.be](https://dmponline.be), pour la création de plans de gestion de données et le respect des obligations liées au RGPD
- Les dépôts institutionnels de données, en production ou en cours de développement (exemple : UCLouvain dispose d'un [Dataverse](https://dataverse.uclouvain.be) depuis 2017)
- Les outils plus spécifiques à certaines disciplines comme le cahier de laboratoire électronique

Prochaines étapes

Ce cadre structurant sera pérennisé même au-delà de la durée du projet RDM en conservant la **communication** entre RDOs et les réunions mensuelles. D'autres projets gérés par cette équipe sont déjà en cours, notamment le projet « HR Excellence in Research – FWB Data Ambassadors community ».

Sensibilisation : conscientiser les utilisateurs des outils

Actions accomplies

Une fois les outils de guidance consolidés et prêts à être utilisés, il convenait de permettre aux utilisateurs de se les approprier. Il ne s'agit pas seulement de pouvoir les utiliser en pratique (ce qui relève plutôt de la formation que de la sensibilisation) mais d'en comprendre pleinement l'utilité et la **plus-value**. On peut citer par exemple la [vidéo](#) produite par UMonS et diffusée largement dans les institutions FWB. Dans le cadre du projet RDM, trois événements de **dissémination** ont permis de sensibiliser un grand nombre de membres de la communauté interuniversitaire aux questions de FAIR Data (cf. Tableau 2). Un point d'honneur a été mis à toujours inviter des orateurs de différentes institutions et disciplines scientifiques, même en-dehors du monde académique.

Toutefois, un problème est rapidement apparu au réseau de RDOs : les initiatives centrales de soutien à la gestion des données manquent de **spécificités disciplinaires**. Plus précisément, il est difficile pour un RDO d'émettre des recommandations immédiatement actionnables pour tous les différents secteurs de recherche scientifique à la fois, tant la **diversité** des habitudes de collecte, d'utilisation, de publication et de réutilisation est grande. Inspirée par l'exemple des pionniers en la matière que sont [Cambridge](#) et [TUDelft](#) [8, 9], une nouvelle approche a émergé dans le projet RDM : **un réseau de sensibilisation par les pairs**. Forts du constat que l'adoption de nouvelles habitudes de travail est plus efficace lorsqu'elles sont transmises entre collaborateurs qui partagent le même environnement de travail [10], les RDOs ont alors décidé de mettre sur pied une **communauté interuniversitaire de Data Ambassadors** (parfois appelés Data Champions dans la littérature). De plus, il était apparu que de nombreux chercheurs des universités francophones faisaient montre d'une certaine expertise dans les questions de gestion, d'analyse, de stockage, d'archivage ou de partage de données, et qu'ils assumaient déjà informellement une mission de soutien et de sensibilisation auprès de leurs collègues immédiats sans pouvoir réellement la **valoriser**. Un titre de Data Ambassador et l'accès à des opportunités de formation et de réseautage, pourraient combler ce manquement.

Après une prise de contact avec les gestionnaires du réseau de Data Champions de TUDelft, l'équipe-projet a proposé le plan d'action suivant :

- Webinaire de présentation du futur réseau de Data Ambassadors sous forme de table ronde avec trois Data Champions de TUDelft (cf. Tableau 2)

- Recrutement des Data Ambassadors : appel ouvert aux candidatures dans toute la communauté universitaire, parfois par l'intermédiaire des autorités (doyens et présidents de départements ou unités de recherche), et prise de contact personnellement avec les chercheurs identifiés comme « déjà » Data Ambassadors *de facto*.
- Prises de contact intra-universitaires : distribution du Guide du Data Ambassador en FWB (Annexe 2), clarification des profils, attentes, motivations et challenges de chacun
- Création et animation d'un espace de réseautage virtuel interuniversitaire (Slack) pour l'échange d'idées, de bonnes pratiques, de questions, de références, ...
- Premières activités interuniversitaires, et ligne et en présentiel

Tableau 2 : Résumé des événements de dissémination et de sensibilisation

Évènement	Thématiques abordées	Note
Webinaire « Data Ambassador Communities » Date : 25/11/21 Organisateur : équipe projet 60 participants	Table ronde avec 3 Data Champions de TUDelft Missions, avantages et challenges du Data Champion Témoignages et exemples d'activités menées par des Data Champions pour leurs pairs	Disponible en replay
Webinaire « Données sensibles et Valorisables » Date : 22/04/21 Organisateur : ULiège 48 participants	Comment reconnaître une situation où s'appliquent les obligations liées au RGPD Notions de propriétés intellectuelles liées aux données particulières et aux données ouvertes Témoignage de mise en place d'un processus efficace de gestion de données qui a permis l'obtention de nombreux brevets (valorisation)	Disponible en replay
Webinaire « Open Data & FAIR principes » Dates : 25 et 26/04/22 Organisateur : ULB 122 participants	Avantages, exigences, obstacles aux partages FAIR et ouvert de données Explication détaillée et pratique (avec exemples) de chaque principe FAIR Data papers	Disponible en replay sur demande Second volet en cours d'organisation

Figure 2 : Capture d'un élément présenté lors du webinaire « Open Data & FAIR Principles » (auteur : Sarah ZAHREDDINE)

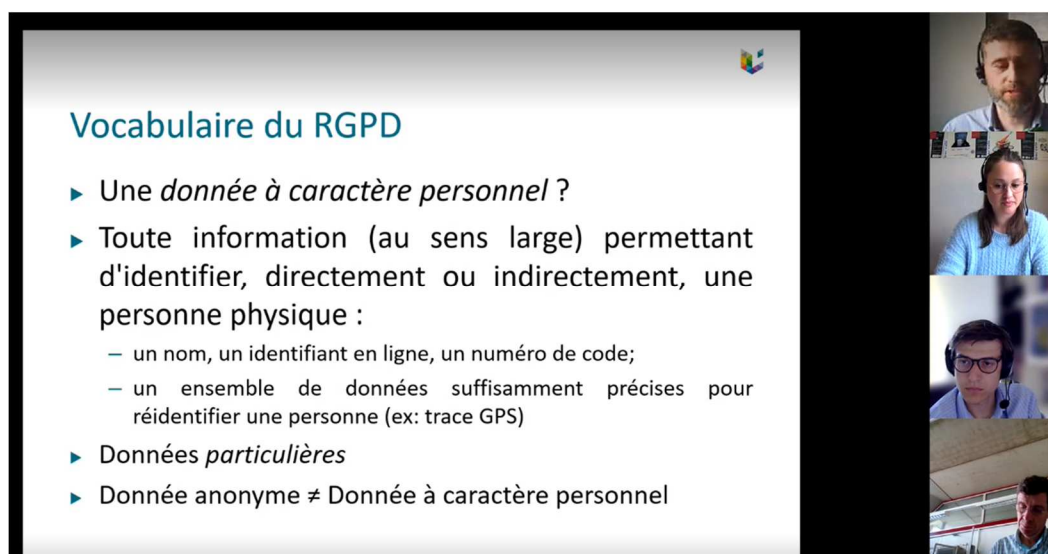


Data repositories

Which repository to select?

- Select a broadly recognized domain-specific repository if available for the specific domain or data type. A general-purpose repository should be selected if no established repository exists for the specific research domain (or for multidisciplinary projects).
- Avoid repositories that require the transfer of rights to the data
- Select a repository that assigns persistent and unique identifiers to the datasets (Digital Object Identifiers, DOIs)
- Select a repository that provides access levels corresponding to the specific needs
- Select a repository that requires entering detailed information about the dataset and uploading documentation (metadata).
- Certified (e.g. CoreTrustSeal <https://www.coretrustseal.org/>)
- Sometimes it may not be possible or not appropriate to deposit data in an external repository, e.g. for legal, ethical, contractual, practical, or other reasons. In such cases, research data selected for preservation will need to be kept in-house.
- Searchable register of existing research data repositories: re3data.org or FAIRsharing.org

Figure 3 : Capture d'un élément présenté lors du webinaire « Données Sensibles et Valorisables » (auteur : Pierre-François PIRLET)



Vocabulaire du RGPD

- ▶ Une *donnée à caractère personnel* ?
- ▶ Toute information (au sens large) permettant d'identifier, directement ou indirectement, une personne physique :
 - un nom, un identifiant en ligne, un numéro de code;
 - un ensemble de données suffisamment précises pour réidentifier une personne (ex: trace GPS)
- ▶ Données *particulières*
- ▶ Donnée anonyme ≠ Donnée à caractère personnel

Le projet RDM a donc jeté les bases de ce réseau, et son **succès** auprès des Facultés est tel qu'il a fait l'objet d'un financement supplémentaire octroyé par la Fédération Wallonie-Bruxelles, sous la coordination de l'ULiège et de l'UCLouvain, pour renforcer ses activités de réseautage et de formations.

Prochaines étapes

Les prochains éléments liés aux Data Ambassadors tombent dans le cadre de ce deuxième projet. Le 24 mai 2022 a eu lieu le kick-off meeting du réseau à l'Université Saint-Louis Bruxelles, qui a recueilli un franc succès auprès des participants : cinq Data Ambassadors ont présenté des bonnes pratiques qu'ils avaient mis en place ou des réussites liées à leur gestion des données. Ces présentations ont recueilli beaucoup d'intérêt et de questions du public, en particulier interdisciplinaires. Une activité de réseautage pour terminer la journée consistait à discuter par petits groupes aléatoires, en un temps limité, sur une thématique précise liée à des difficultés communes en gestion des données, avec l'aide d'un modérateur. Il est alors apparu à quel point les chercheurs se saisissent de l'opportunité d'interactions interdisciplinaires et interinstitutionnelles sur cette thématique, pour mieux consolider leurs habitudes de travail, renforçant encore la motivation des RDOs à continuer d'entretenir ces liens. Est également prévue au calendrier une demi-journée de formation en ligne sur des thématiques générales de gestion des données auprès du [Data Curation Center](#) (UK).

Les premières phases de de travail font l'objet d'une **présentation de poster lors de la conférence LIBER 2022**, rendez-vous périodique des bibliothèques universitaires européennes. Dans le cadre de ce poster, une **enquête** sera conduite auprès des Data Ambassadors pour étudier leurs motivations, leurs attentes, et leur auto-évaluation de leur niveau d'expertise en gestion des données avant et après les activités du réseau. Le défi le plus immédiat consiste à relancer des activités en présentiel, pour lesquelles l'engagement reste plus difficile au vu des nouvelles habitudes de travail à distance acquises pendant la pandémie de COVID-19.

En complément des activités de sensibilisation menées par les RDOs, la mission centrale du réseau de Data Ambassadors est la **sensibilisation entre pairs**. Ainsi, le focus sera placé sur cette communauté et sur les activités que les Data Ambassadors **eux-mêmes** vont organiser, encouragés et soutenus par les universités. Lors de leur engagement, les Data Ambassadors ont été informés qu'*a minima*, une

activité de sensibilisation était attendue de leur part d'ici le printemps 2023 (soit un an après le kick-off meeting du réseau). Il pourra s'agir de séminaires, documents, newsletters, tables rondes, vidéos, ou tout format qui leur convient.

Formation : connaissances et autonomie

Actions accomplies

Le développement de la carrière du chercheur est une priorité des universités de la Fédération Wallonie-Bruxelles. L'un de ses aspects est leur **formation continue**, en compétences soit thématiques soit transversales. Les questions de gestion de données de la recherche sont directement pertinentes pour l'activité scientifique, mais peuvent aussi être transférées en-dehors du monde académique au besoin. Proposer des formations à ces questions est donc **essentiel** pour les universités.

Ces formations apportent des compléments **pratiques** aux thématiques de sensibilisation, via les outils développés dans l'axe de travail « guidance » mais aussi via des outils extérieurs ou des habitudes de travail à s'approprier. Pour ces actions, un excellent point de départ a été trouvé dans un livrable d'un projet précédent : le [portail web du Research Data Management](#) créé lors d'un projet mené par UCLouvain et la BICFB. Ces ressources, majoritairement externes, ont servi de base pour construire, chacun dans son institution mais en concertation, des **ressources internes** et séances de formation régulières. Elles sont organisées au sein de chaque institution, pour être le mieux adaptées possible à la réalité de chaque chercheur. Elles sont conduites le plus souvent par les RDOs mais en collaboration avec les autres services clés dans la gestion des données, tels que les services juridiques, informatiques, ou bien sûr les bibliothèques. Un document de travail pour inventorier les diverses formations proposées a été dressé, qui a encore renforcé l'évidence du manque de « percolation » des recommandations générales dans les diverses disciplines scientifiques, démontrant encore la **pertinence** du réseau de Data Ambassadors.

Sur la période de mars 2021 à janvier 2022, au total, **29 séances** de formations généralistes ont été organisées et ont rassemblé un total de **819 chercheurs**. Des formations plus spécifiques ont été organisées également. Par exemple, sur la même période écoulée à l'ULiège :

- Une formation-exercice sur l'outil DMPonline.be a recueilli **39 participants sur 3 séances**
- **106 doctorants et chercheurs** ont suivi au moins une formation complémentaire en **statistiques**
- **34 doctorants et chercheurs** ont participé à une séance de formation sur **les bases de l'Open Access et l'Open Science** en général.

Des séances supplémentaires de formation ou d'aide individuelle aux outils de la gestion de données sont délivrées sur demande, par exemple lorsqu'un chercheur en fait la demande dans le cadre d'un projet de recherche, ou au sein d'une unité de recherche.

Tableau 3 : liste non exhaustive des thématiques abordées lors des séances de formation généralistes délivrées par les RDOs

FAIR Data	Reproducibility	Repositories	Format
Open Data	Visibility	Specific regulations	Quality
Data Management Plans	Publication	GDPR	P-hacking
Sharing	Backup	Property	Integrity
Re-use	Security	Licenses	Value
Metadata	Storage	Ownership	Standards

Prochaines étapes

D'une part, il conviendra pour les RDOs de continuer de se former, notamment aux nouveaux outils liés au *European Open Science Cloud* et aux **nouvelles exigences** des programmes de financement belges et européens, pour continuer de **garantir** aux chercheurs les outils et compétences nécessaires en gestion des données dans le montage de tels projets. Les formations généralistes sont vouées à être pérennisées d'une part, et complémentées, d'autre part, par des formations plus spécifiques et plus approfondies, par exemple sur les réglementations applicables à certains types de données, ou sur la gestion de très grands volumes de données, ou encore sur l'intégrité scientifique dans les tests statistiques. Ces approfondissements seront construits en mobilisant les ressources clés de la gestion de données, comme les services juridiques, informatiques, ou scientifiques lorsque c'est nécessaire. Enfin, le réseau de Data Ambassadors donne également aux RDOs l'accès à cette expertise spécifique **disciplinaire** dont ils peuvent parfois manquer. À titre d'exemple, une collaboration entre les Data Ambassadors des sciences humaines et sociales et la RDO de l'UCLouvain, va permettre d'élaborer une formation spécifique à la bonne gestion des données qualitatives.

Perspectives et conclusion

L'approche **interuniversitaire** de ce projet a apporté une véritable **plus-value**, d'une part en optimisant le travail et en capitalisant sur les ressources existantes, et d'autre part, en **accéléralant** et accroissant directement la portée de chaque action. Le cadre structurant du réseau de RDOs restera d'actualité pour conserver cette centralisation des initiatives liées à la gestion des données en FWB. Le projet RDM donne lieu aux perspectives suivantes :

Guidance

- Suivi et mises à jour des guidances sur l'outil DMPonline.be
- Suivi de la mise en place des répertoires de dépôts de données lorsque c'est nécessaire
- Suivi des outils liés aux dynamiques belges et européennes (par exemple EOSC)

Sensibilisation

- Accroissement et pérennisation du réseau FWB de Data Ambassadors (via la suite du projet RDM)
- Concertation sur le développement de nouveaux supports de sensibilisation (flyers, pages web, ...)

La proposition d'ouvrir les activités des Data Ambassadors aux établissements d'enseignement supérieur est également à l'étude.

Formation

- Accroissement de la formation des RDOs, individuellement et collectivement
- Accroissement de l'offre de formations disponibles aux chercheurs :
 - o Pérennisation des formations généralistes
 - o Compléments à ces formations pour approfondir les thématiques spécialisées
 - o Tutoriels et formations appliquées aux futurs outils déployés le cas échéant

De manière générale, le projet RDM a permis non seulement de structurer et renforcer l'intérêt des communautés universitaires de la Fédération Wallonie-Bruxelles pour les bonnes pratiques de gestion des données, mais aussi de démontrer la plus-value liée à l'approche interuniversitaire.

Remerciements

Les auteurs remercient la Fédération Wallonie-Bruxelles pour le soutien qu'elle a témoigné à ce projet. Elles remercient également les contributeurs et relecteurs de ce document, en précisant que toute erreur qui s'y trouverait encore devrait leur être entièrement attribuée. Parmi ces relecteurs, un remerciement particulier est adressé à l'équipe en charge du projet composée d'Adeline GRARD, Delphine LOBET, Céline THILLOU, Françoise VANDOOREN, Sarah ZAHREDDINE, Dodji AMOUZOU et Bernard DETREMBLEUR.

Bibliographie

- [1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- [2] Ioannidis JPA. Contradicted and Initially Stronger Effects in Highly Cited Clinical Research. *JAMA*. 2005;294(2):218–228. doi:10.1001/jama.294.2.218
- [3] Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (2016). <https://doi.org/10.1038/533452a>
- [4] Mark Otto Baerlocher, Jeremy O'Brien, Marshall Newton, Tina Gautam, Jason Noble, Data integrity, reliability and fraud in medical research, *European Journal of Internal Medicine*, Volume 21, Issue 1, 2010, Pages 40-45, ISSN 0953-6205, <https://doi.org/10.1016/j.ejim.2009.11.002>.
- [5] Higman, Rosie, Daniel Bangert, and Sarah Jones. 2019. "Three Camps, One Destination: The Intersections of Research Data Management, FAIR and Open". *Insights* 32 (1): 18. DOI: <http://doi.org/10.1629/uksg.468>
- [6] Nelson, Nicole C. Understand the real reasons reproducibility reform fails, *Nature* 600, 191 (2021), doi: <https://doi.org/10.1038/d41586-021-03617-w>
- [7] Nuzzo, R. Scientific method: Statistical errors. *Nature* 506, 150–152 (2014). <https://doi.org/10.1038/506150a>
- [8] Connie Clare. (2019). How to build a community of Data Champions: Six Steps to Success. Zenodo. <https://doi.org/10.5281/zenodo.3383814>
- [9] Savage, JL and Cadwallader, L. 2019. Establishing, Developing, and Sustaining a Community of Data Champions. *Data Science Journal*, 18: 23, pp. 1–8. DOI: <https://doi.org/10.5334/dsj-2019-023>
- [10] Clare, Connie. (2019) The Real World of Research Data. Zenodo. <https://doi.org/10.5281/zenodo.3584373>

Annexe 1 - Résumé des livrables

	Intitulé	Statut	Notes
1 et 2	Modèle de DMP, sous forme de document ou intégré dans DMPonline, à disposition des chercheurs Intégration de l'obligation du registre RGPD dans le modèle de DMP et/ou dans DMPonline		Disponible dans toutes les institutions, modèles H2020 et/ou institutionnels Intégration RGPD en cours ou complète.
3	Cahier de laboratoire électronique : version au moins pilote, à l'échelle institutionnelle OU à l'échelle des UR		Utilisation du cahier LIEU toujours répandue mais projet au moins pilote de cahier de laboratoire électronique dans certaines institutions
4	Répertoire institutionnel (DataVerse ou autre) de dépôt des (méta)données avec possibilité d'ouverture d'accès aux données et documentation associée : au moins version pilote		Sous l'impulsion du projet MQTOS, en production dans certaines institutions, en développement dans les autres, en concertation (partage d'expériences)
5	Formaliser les rôles de Research Data Officier au sein de chaque institution		Disponible dans toutes les institutions
6	Formaliser les rôles de <i>Data Champion</i> au sein de chaque institution		Création du Réseau de Data Ambassadors avec Handbook/Guide du Data Ambassador (en français et en anglais), Slack permettant l'échange interdisciplinaire et interuniversitaire
7	Inventaire et recommandations de formations disponibles pour compléter les compétences des RDO		Exemples de formations suivies : DU SDM Montpellier Inventaire des formations mises en œuvre dans les différentes universités, pour partage d'expérience et des supports de formation
8	Critères de sélection de <i>data repositories</i>		Disponibles ici Diffusés lors du webinaire "Open Data & FAIR Principles"
9	Mise en réseau des RDO de chaque institution, notamment par des réunions périodiques		Réunions mensuelles depuis 4 mars 2021 Réseau de Data Ambassadors (<i>community management</i>)
10	Définition des indicateurs et métriques pour suivi du « climat » RDM de la FWB		Collaboration, via GT OS du CRef, à la rédaction de la checklist LERU pour ses aspects RDM
11	Catégories disponibles dans la plateforme de dépôt institutionnel de <i>data papers / reviews / replication papers</i>		Collaboration à la réflexion conjointe avec les bibliothèques institutionnelles
12 à 18	Axe de sensibilisation		Cf. « Événements de dissémination » Réutilisation de matériel du projet BICFB Réseau de Data Ambassadors, en particulier, création d'un espace Slack interuniversitaire
19 à 23	Axe de formation		Séances de formation généralistes organisées dans chaque institution. Sur la période écoulée (04/03/2021 – 04/05/2022) : - ULiège : 7 séances, 117 participants au total - UCLouvain : 12 séances, 237 participants au total - ULB : 10 séances (465 participants au total) - UMon : vidéo diffusée auprès des chercheurs - UNamur : Aide sur demande - USL-B : aide sur demande Projet de suivi : pérenniser + étendre à d'autres thématiques (données qualitatives) pour renforcer leur caractère disciplinaire et interuniversitaire

Annexe 2 - Guide du Data Ambassador en FWB

Guide du Data Ambassador en FWB

Décembre 2021

Vous avez décidé de devenir un Data Ambassador. Grâce à votre engagement, vous allez **promouvoir les bonnes pratiques de gestion des données de la recherche** dans votre discipline scientifique et, avec le soutien de votre Université, contribuer à leur adoption au sein de la FWB.

Pour mener à bien cette mission de Data Ambassador, vous pourrez compter sur l'appui d'une communauté interuniversitaire de personnes engagées dans cette démarche. Les Data Ambassadors des six universités de la FWB forment un **réseau collaboratif** au sein duquel vous serez amenés à partager vos idées, vos interrogations et vos initiatives. Même si chaque Data Ambassador œuvre dans son environnement immédiat (département, faculté, unité de recherche, école doctorale...) cette démarche collective assurera une dynamique durable et engagée avec une vision commune : une recherche plus ouverte en FWB.

En plus de cette communauté de Data Ambassadors, votre institution mettra tout en œuvre pour **vous aider** autant que possible dans vos initiatives, par exemple en vous proposant une aide pratique dans l'organisation de vos activités ou du matériel de sensibilisation que vous serez libre de réutiliser. Votre rôle de Data Ambassador vous donne également accès à des opportunités uniques en termes de réseautage et de formations. Par la visibilité que vous gagnerez en exerçant votre mission, votre université vous aidera à la valoriser auprès de votre communauté scientifique et professionnelle.

Science ouverte, FAIR data et cycle des données

La gestion et la planification des données de la recherche fait partie des exigences de plus en plus d'organismes de financement. Au-delà de cette obligation, le Research Data Management (RDM) présente **de très nombreux avantages** pour le travail d'un chercheur. Enfin, le RDM facilite la démarche d'ouverture des données de la recherche, et même d'**Open Science** en général.

L'Open Science désigne « *une approche de la démarche scientifique fondée sur le travail coopératif et sur de nouveaux modes de diffusion des connaissances, améliorant l'accessibilité et la réutilisation des résultats de la recherche en utilisant les technologies numériques et les nouveaux outils de collaboration*¹. Il s'agit de partager les résultats de la recherche le plus tôt possible dans le processus de recherche, notamment les données de la recherche (Open Data), les outils numériques (Open Source), les publications scientifiques (Open Access), et de manière à les rendre reproductibles et réutilisables. Ces pratiques permettent de rendre la recherche plus collaborative, plus transparente, plus fiable et plus efficace ». L'Open Science donne véritablement vie à un résultat scientifique, en lui permettant de nourrir d'autres recherches.

Dans certains cas, comme par exemple lorsque les données revêtent un caractère confidentiel (données sensibles, brevetables), ouvrir l'accès aux données peut s'avérer impossible.

¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018H0790&from=EN>

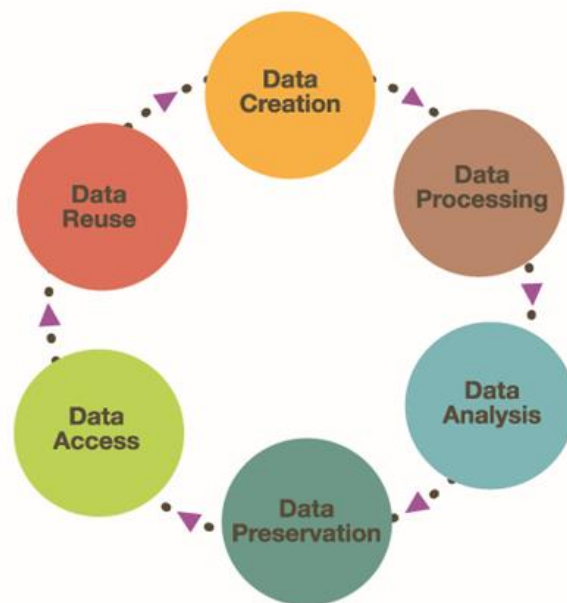
Néanmoins les données peuvent être préparées pour une réutilisation ultérieure en appliquant les **principes FAIR**. L'objectif est d'ouvrir l'accès aux données autant que possible et d'en restreindre l'accès autant que nécessaire.

FAIR est un acronyme qui réfère aux concepts suivants.

- *Findable*, ou faciles à trouver : les données doivent être trouvables par la communauté scientifique. Pour ce faire, elles sont décrites par des **métadonnées**. Les métadonnées sont des informations sur l'origine et le contenu d'un jeu de données (auteur, date, mots-clés, ...). Le plus souvent, elles sont compréhensibles aussi bien par les humains que par les machines. Pour être faciles à trouver, les données, lorsqu'elles sont publiées, peuvent être associées à un identifiant pérenne, de type DOI ou autre. De cette manière, tous les travaux qui en découlent pourront s'y référer. Pour obtenir des DOI ou une structure de métadonnées appropriée, lorsque les spécificités du projet le permettent, les données peuvent être déposées dans un **dépôt de données** (*data repository*) en ligne, comme Zenodo, OSF ou Figshare, ou le dépôt de données de votre université, s'il en existe un. De nombreux dépôts de données dédiés à un secteur ou une discipline scientifique existent également, et vous en connaissez sans doute !
- *Accessible* ou accessibles : les données doivent être mises à disposition et être accessibles de manière durable, et leur **pérennité** assurée pour une conservation à très long terme. Elles sont **aussi ouvertes que possible**, en respectant toutefois les spécificités du cadre de la recherche qui les a générées. Dans la plupart des cas, si les données ne peuvent être partagées, les métadonnées le peuvent à tout le moins
- *Interoperable* ou interopérables : les données doivent pouvoir être **échangées, communiquées, interprétées et utilisées** par divers logiciels et applications, dans un langage clair et compatible avec un maximum de standards. Il s'agit par exemple de choisir des formats de fichiers lisibles par des logiciels libres, ou des formats de données très utilisés dans votre discipline, ou de rédiger la documentation associée en anglais.
- *Reusable* ou réutilisables : les données sont suffisamment décrites et partagées avec le moins de **barrières** possibles (licences ouvertes, formats portables, ...) pour être réutilisées par le plus grand nombre. Les métadonnées sont détaillées, disponibles, compréhensibles, et alignées avec les standards de la communauté concernée.

L'application des principes FAIR est la clé d'une recherche plus ouverte, même lorsque l'ouverture de l'accès aux données n'est pas possible. Le respect de ces principes est de plus en plus souvent exigé par les organismes de financement en Belgique et en Europe, qui requièrent la rédaction d'un **Plan de Gestion des Données** (*Data Management Plan*) pour s'assurer de la concordance entre le projet de recherche et ces principes FAIR. Quoi qu'il en soit, ces principes se traduisent en **pratiques concrètes** dans le travail de gestion des données. Cette dernière englobe la totalité des processus que les données subissent, incluant mais ne se limitant pas à la planification, la collecte, le stockage, l'analyse, la sauvegarde, la documentation, le référencement, la publication et le partage.

À chacune des étapes du **cycle de vie des données de la recherche**, de bonnes habitudes peuvent être mises en place pour appliquer les principes FAIR ou ouvrir l'accès aux données.



Quelques exemples :

- Rédiger systématiquement un plan de gestion des données, même lorsque le financement ne l'exige pas
- Anticiper les **réglementations** auxquelles les données peuvent être soumises (données sensibles, données biologiques, données humaines, données confidentielles, données liées à un brevet...)
- Conserver les données originales de manière pérenne pour assurer la **traçabilité** des résultats obtenus
- Assurer un stockage **sécurisé** de ses données ainsi que des backups
- Documenter minutieusement chaque étape de la **méthodologie** de recherche déployée
- Choisir des **formats** d'encodage les plus portables possibles
- **Documenter** les jeux de données à l'aide de métadonnées ou de fichiers auxiliaires (auteur, date et méthode de collecte, source, structure, description des variables...)

Vos missions et vos opportunités

En tant que Data Ambassador, vous aidez les chercheurs de votre environnement de travail (équipe, département, faculté, unité de recherche, école doctorale...) à développer ces bonnes pratiques de gestion de leurs données. Vos interactions avec les chercheurs peuvent revêtir plusieurs formes.

Sensibilisation : il vous appartient d'organiser à votre manière des activités de sensibilisation aux bonnes pratiques de la gestion des données. Ces activités doivent avoir pour but de traduire de manière spécifique à votre réalité scientifique les recommandations générales de la gestion des données et les principes FAIR. Sur base du matériel de sensibilisation fourni par le réseau, l'institution, ou en construisant votre propre support, vous diffusez auprès des chercheurs les questions à se poser et les bonnes habitudes à adopter au sujet de la gestion des données. Vous pouvez organiser des séminaires généraux ou thématiques, construire une FAQ, contacter vos collègues par e-mail... La parole est à vous. Votre institution vous soutient dans cette démarche autant que possible. N'hésitez pas à demander de l'aide à votre gestionnaire de réseau, votre Research Data Officier (RDO) ou votre Administration de la Recherche.

Workshops & Presentations	Other Activities
File management	Open Data FAQ
Data sharing	Bite-size emails
Avoiding data disasters	Training needs analysis
Writing data management plans	1:1 RDM and data analysis support
Repositories	Data Policy
Code management	Data Audit
GitHub introduction	Outreach materials
'Bring your own data' workshop	Electronic Lab Notebook trial

Exemples d'autres activités lancées par les Data Ambassadors de l'Université de Cambridge (Source : Savage, JL and Cadwallader, L. 2019. Establishing, Developing, and Sustaining a Community of Data Ambassadors. Data Science Journal, 18: 23, pp. 1–8. DOI: <https://doi.org/10.5334/dsj-2019-023>)

Soutien : il pourrait arriver qu'un collègue vous demande de l'aide individuelle. Bien que votre mission revête un rôle de conseil, vous vous appuyez sur de nombreux partenaires : votre gestionnaire de réseau, votre RDO, votre Administration de la Recherche, le réseau de Data Ambassadors, ou tout autre organisme de soutien à la Recherche. Ainsi, vous pouvez adopter le réflexe de rediriger les demandes vers les personnes ressources adéquates. Être un Data Ambassador ne vous oblige nullement à vous mettre au service de quiconque. Votre rôle est celui d'un conseiller, vous n'exécutez pas de travail pour vos collègues.

Dialogue avec l'institution : en plus du rôle que vous occupez dans votre environnement de travail proche, vous interagissez régulièrement avec le gestionnaire institutionnel du réseau des Data Ambassadors. Lorsque c'est nécessaire, vous relayez vers lui les demandes que vous recevez, surtout si elles sont récurrentes. Vous lui rapportez les problèmes courants rencontrés avec les outils et services institutionnels, les besoins collectifs en formation ou en ressources, et si vous le souhaitez, vous lui proposez des améliorations.

Parallèlement, votre institution peut compter sur vous pour une expertise spécifique dans les décisions centrales relatives au RDM lorsque c'est nécessaire.

Opportunités : vous participez à un réseau interuniversitaire. Vous avez accès à des **événements** et **espaces** de partage enrichis par des collègues aux profils variés et aux compétences complémentaires. Vous agrandissez ainsi votre réseau professionnel. Grâce à cette initiative, vous avez accès à des formations ciblées.

Combien de temps devez-vous consacrer aux missions de Data Ambassadors ? Autant que vous le souhaitez. Vous exercez cette mission de manière volontaire et en plus de votre charge de travail existante, vous disposez donc d'une grande liberté de flexibilité dans la manière dont vous la menez à bien. **A minima, votre institution aimerait organiser avec vous au moins une activité de sensibilisation par an, quelle qu'elle soit.**

Puisque votre mission de Data Ambassador est volontaire, elle mérite d'être **reconnue** et valorisée. À cet effet, votre institution favorise votre visibilité aussi bien en interne qu'en externe : par exemple, votre nom et/ou votre profil pourra apparaître sur les pages web du réseau de Data Ambassadors, et votre doyen sera informé de votre démarche le cas échéant. Cette expérience constitue une plus-value certaine à votre curriculum, et les compétences que vous développerez sont très recherchées dans le monde du travail.

Le réseau de Data Ambassadors de la FWB

Personnes de contact

Gestionnaire globale du réseau de Data Ambassadors	Judith Biernaux jbiernaux@uliege.be
Gestionnaire UCLouvain	Adeline Grard adeline.grard@uclouvain.be
Gestionnaire UMon	Céline Thillou celine.thillou@umons.ac.be
Gestionnaire UNamur	Bernard Detrembleur bernard.detrembleur@unamur.be
Gestionnaire ULB	Sarah Zahreddine Sarah.Zahreddine@ulb.be
Gestionnaire ULiège	Judith Biernaux jbiernaux@uliege.be
Gestionnaire USL-B	Delphine Lobet delphine.lobet@usaintlouis.be