



LE PROJET DATA LIFE-CYCLE MANAGEMENT (DLCM) EN SUISSE : UNE GESTION DES DONNÉES DE LA RECHERCHE PENSÉE POUR SES UTILISATEURS

[Pierre Yves Burgi](#), [Basma Makhoulf Shabou](#)

A.D.B.S. | « I2D - Information, données & documents »

2021/2 n° 2 | pages 87 à 95

ISSN 2428-2111

DOI 10.3917/i2d.212.0087

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-i2d-information-donnees-et-documents-2021-2-page-87.htm>

Distribution électronique Cairn.info pour A.D.B.S..

© A.D.B.S.. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Mots clés : Gestion des données de la recherche, Open Science, Recherche, Suisse

LE PROJET DATA LIFE-CYCLE MANAGEMENT (DLCM) EN SUISSE : UNE GESTION DES DONNÉES DE LA RECHERCHE PENSÉE POUR SES UTILISATEURS

Lancé en août 2015 dans le cadre du programme national *Informations scientifique* visant à poser les bases de l'*Open Science* en Suisse, et cofinancé par swissuniversities, le projet *Data Life-Cycle Management (DLCM)* fournit aux chercheurs des accès et des partages aisés aux publications scientifiques et aux données de recherche.



Dr. Pierre Yves BURGHI

Université de Genève, co-directeur du projet DLCM, Université de Genève.

+ pierre-yves.burgi@unige.ch

Prof. Dr. Basma MAKHLOUF SHABOU

Co-directrice du projet DLCM, HEG/HES-SO, est responsable du MSc IS, Haute école de gestion de Genève (HEG), Haute école spécialisée de Suisse occidentale (HES-SO).

+ basma.makhlouf-shabou@hesge.ch



OPEN SCIENCE

Le projet *Data Life-Cycle Management* (DLCM) a été lancé officiellement en août 2015 dans le cadre du programme national « Information Scientifique » visant à poser les bases de l'*Open Science* en Suisse, par des accès et partages aussi aisés que possible aux publications scientifiques et aux données de recherche.

L'objectif premier de ce projet est de mettre en place des services destinés principalement aux chercheurs pour leurs besoins les plus importants tout au long du cycle de vie de leurs données.

L'identification de ces besoins s'est basée sur une étude de la littérature, ainsi que des entretiens semi-structurés, ce qui a conduit à répertorier les principaux scénarios d'utilisation des données de recherche, ainsi que des services de base :

- un point d'accès et de contact pour obtenir des informations,
- de la formation, et des conseils personnalisés,
- des solutions de gestion active des données, notamment des options de stockage, de conservation à long terme et de publication selon des normes internationales.

Le projet DLCM a été réalisé en deux phases.

2015-2018 : PHASE 1 DU PROJET DLCM, PRÉÉTUDES, PROTOTYPES

Cette première phase du projet, qui s'est concentrée sur des pré-études et la conception de prototypes touchant à la gestion active des données et à leur archivage, a bénéficié de l'expertise de spécialistes de l'information scientifique, d'informaticiens et de services à la recherche des hautes écoles suisses^[1]. Lors des consultations auprès des chercheurs couvrant une trentaine de disciplines^[2], un point mentionné régulièrement concerne le manque de réponse adéquate à la question de ce qui pourrait et devrait arriver aux données de recherche après la fin d'un projet et / ou une fois que la publication des résultats dans un journal devient effective. La citation de données dans la littérature, ou de nouvelles méthodes d'évaluation qui valoriseraient le travail de recherche effectué sur les données, comme les *data papers*^[3], sont aussi souvent mentionnées.



Cette phase, « orientée chercheur », a permis d'identifier les composants de base, contribuant au développement d'un référentiel générique de données, adapté à tous les chercheurs et toute institution. Au moment de leur lancement initial au terme de cette première phase, l'ensemble des services de base de la solution DLCM comprenait^[4] :

- Un formulaire type pour le *Data Management Plan* (DMP), avec accompagnement et formation à l'appui (assurés par les bibliothèques) ;
- Un *DLCM Policy Tool* pour faciliter et soutenir le processus d'élaboration d'une politique de gestion des données de recherche au niveau institutionnel ;
- Un tableau comparatif des solutions de gestion active des données^[5] ;
- Une cartographie des experts sur différents aspects de la gestion des données de la recherche ;
- Une intégration des enseignements sur les données de recherche au sein du Master en sciences de l'information à la Haute école de gestion de Genève ;
- Un concept de service de consultation établi en réseau pour répondre aux questions des chercheurs et des professionnels de l'information ;
- Un programme de formation annuel en ligne et sur site proposé en plusieurs langues dans différents cantons suisses ;
- Une solution de préservation long terme des données.

2019-2020 : PHASE 2 DU PROJET DLCM, PROJETS PILOTES

Cette phase, correspondant au passage des prototypes aux services nationaux, a impliqué une dizaine de projets pilotes^[6], centrés sur les chercheurs, pour affiner l'ensemble des services de base avec des fonctionnalités supplémentaires. Une année d'exploitation de la solution DLCM et de soutien aux utilisateurs au niveau des hautes écoles du canton de Genève a permis de mieux comprendre le contexte unique de chaque institution et chercheur^[7], et a influencé la sélection et la priorisation des développements qui répondent au mieux à leurs exigences^[8]. Parmi ces développements, nous pouvons citer trois catégories : (1) les fonctionnalités justifiant l'assurance et le contrôle de la qualité, (2) les fonctionnalités apportant un impact grâce à une contextualisation supplémentaire, par exemple au travers de politiques d'archivage per-



sonnalisables, des possibilités d'intégration des instruments de laboratoire, des outils de visualisation et d'analyse, etc., et (3) les fonctionnalités permettant la sécurité tout en assurant la visibilité, la réutilisation et la reproductibilité des résultats de recherche.

LA TECHNOLOGIE DLCM

Basée sur une architecture modulaire (Figure 1), conforme au modèle de référence OAIS (ISO 14721)^[9], aux principes FAIR^[10], et aux normes internationales^[11], la solution de préservation long terme se caractérise par :

- Des connecteurs à plusieurs systèmes de stockage d'archives tels que le système de fichiers, le protocole d'Amazon S3 et les bandes magnétiques.
- Un mécanisme d'identifiant persistant basé sur l'identifiant d'objet numérique (DOI), avec les métadonnées minimales requises basées sur le schéma ouvert de métadonnées descriptives DataCite^[12].
- Des mesures de sécurité telles que la redondance, la séparation géographique des copies, les checksums et mécanismes d'audit - par défaut, les données restent en Suisse, mais les clients peuvent configurer des nœuds d'archivage chez tout autre fournisseur de stockage.
- La gestion des licences, avec un choix par défaut de licences Creative Commons^[13] et d'autres conditions personnalisables.
- La possibilité d'une période d'embargo.
- L'ajout automatique de métadonnées en format PREMIS^[14].
- L'indexation et le moissonnage par le protocole OAI-PMH^[15].
- Un support multilingue (français, allemand et anglais).

Le principal avantage concurrentiel de la technologie développée dans le projet DLCM provient de son architecture modulaire et distribuée, ainsi que de sa stricte conformité au modèle de référence OAIS de la norme ISO 14721 (2012) (Figure 1). Aux trois entités standards de cette norme (SIP, AIP, DIP), nous avons ajouté un module de pré-ingestion. Ce module permet une grande flexibilité dans la gestion des données en offrant aux chercheurs la possibilité de manipuler les jeux de données avant leur soumission finale. La pré-ingestion intervient donc après la phase de travail actif sur les données, mais avant la phase d'archivage, qui empêche toute modification ultérieure. Toute modification des données après l'archivage impliquerait soit un nouveau jeu de données archivés, soit, lorsque cela est autorisé, une nouvelle version du jeu



de données original. Une telle gestion des versions n'est pas autorisée ici, mais peut être remplacée par la notion de collections, qui pourrait regrouper plusieurs versions successives du jeu de données original.

Des rôles prédéfinis permettent de définir différents groupes d'utilisateurs, par exemple en donnant aux co-auteurs le droit d'éditer l'ensemble des données tout en restreignant la visualisation à un groupe spécifique d'utilisateurs (par exemple, les visiteurs). Les rôles permettent également de mettre en place un contrôle de qualité, effectué soit par les gestionnaires, les responsables ou les approbateurs par le biais d'un flux de travail. L'activation d'un tel flux de travail reste facultative et n'aurait pas de sens si l'institution/département/laboratoire n'a pas de stratégie de qualité des données.

Une autre caractéristique spécifique à l'architecture DLCM est la notion d'unités organisationnelles. Ces dernières permettent d'organiser les ensembles de données au sein d'unités dont la granularité peut être fixée au niveau d'un projet, d'un laboratoire, d'un département ou d'une institution. Une telle organisation peut être un instrument puissant pour suivre les indicateurs clés, et s'avère également pratique pour structurer logiquement un grand nombre d'ensembles de données.

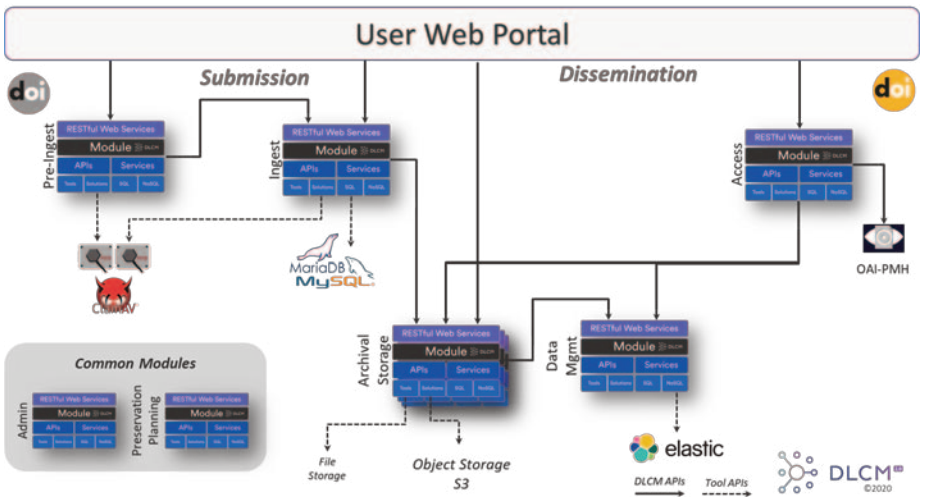


FIGURE 1: ARCHITECTURE MODULAIRE DLCM POUR LA PRÉSERVATION LONG TERME DES DONNÉES

Toutes les fonctionnalités (dépôt, téléchargement, recherche, etc.) sont accessibles à partir de tout environnement capable d'activer des services web. Par exemple, des connecteurs *Jupyter Notebook*^[16] permettent de rechercher et



d'extraire des données, pour ensuite les traiter sur la base d'une variété de langages. Pour de grands volumes de données (plus de 100 Go) ou un nombre élevé de fichiers (plusieurs milliers), nous fournissons une assistance aux chercheurs afin que l'ingestion soit automatisée par lots sur la base de « scripts » *ad hoc*.

SERVICES COMPLÉMENTAIRES

Divers services sont proposés en compléments à la pérennisation. Ils se focalisent sur trois axes: la formation, la consultation et les enseignements en ligne et en présentiel dans le domaine de la gestion des données de la recherche.

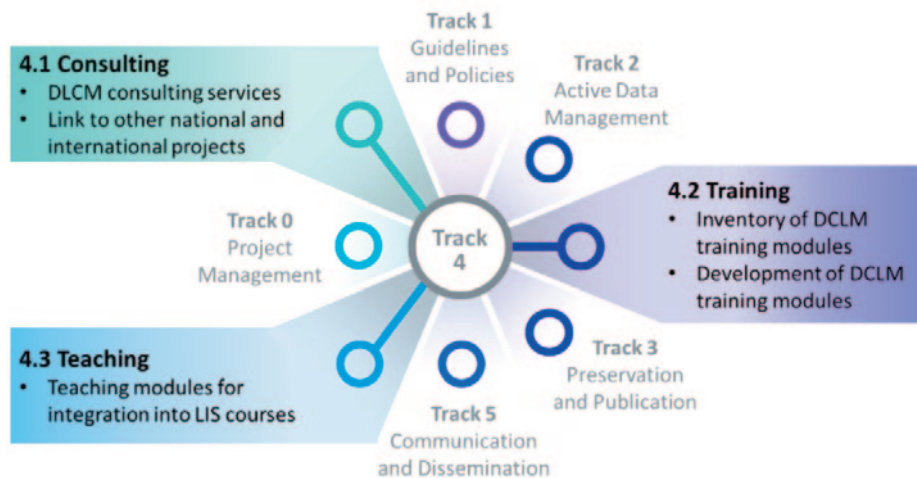


FIGURE 2: OBJECTIFS, LIVRABLES ET COLLABORATION DE LA DIMENSION 4 DANS LE PROJET DLCM © DLCM – CC BY-NC-SA

FORMATION

Au cours des deux dernières années, plus d'une centaine de personnes ont bénéficié de différentes sessions de formation, pour les chercheurs principalement. Ces sessions ont été livrées, en distanciel et présentiel en anglais, en français et en allemand, et ce, dans différentes régions de Suisse (Genève, Fribourg, Lausanne, Berne et Zurich). Les thématiques couvrent essentiellement les plans de gestion de données, les questions du stockage, d'archivage pérenne, les aspects légaux réglementaires, les données qualitatives et quantitatives, et leur plan de traitement.

Un MOOC est également en cours de réalisation, grâce à la collaboration d'experts renommés.



CONSULTATION

Un concept de conseil aux chercheurs a été établi depuis la fin 2018 afin de compléter le service de pérennisation des données de recherche. Ce service est proposé à toute la communauté scientifique suisse et consiste à répondre aux requêtes reçues via notre adresse DLCM ou encore relayées via nos partenaires et contacts dans les institutions académiques.

Les requêtes sont traitées dans un temps relativement court: deux jours ouvrables. La plupart des questions concernent des aspects liés au plan de gestion de données et les choix que les chercheurs doivent y annoncer pour mieux répondre aux exigences des bailleurs de fonds et des agences de financement.

Deux niveaux sont possibles: la consultation de niveau 1 propose des réponses directement aux chercheurs sur des aspects de base, et la consultation de niveau 2, fournit le contact d'un spécialiste dès qu'il s'agit d'un aspect technique ou trop spécifique.



FIGURE 3: CONCEPT DE CONSULTATION SELON LE PROJET DLCM

DLCM a pu construire un réseau étendu de spécialistes des données de recherche dans diverses institutions académiques de Suisse. Ce service de consultation est proposé, par mail et/ou par téléphone, en anglais et dans trois langues officielles de Suisse.



CONCLUSION

Le projet DLCM a développé une technologie conforme aux normes et aux exigences légales, réglementaires et éthiques aux niveaux suisse et international. Un réseau d'experts et de connaissances a été généré également durant les différentes phases et étapes de ce projet.

Deux plateformes de pérennisation ont bénéficié des réalisations du projet DLCM: la première nommée Yareta^[17] prend en charge les données des chercheurs du Canton de Genève et la deuxième nommée OLOS^[18] vise tous les chercheurs de Suisse. Cette dernière est gérée par l'Association OLOS, qui prendra la relève en matière de gestion et de pérennisation des données de la recherche, et ce, aux niveaux national et international.

RÉFÉRENCES

- [1] **BLUMER E., BURGI, P.-Y.** (2015). Data Life-Cycle Management Project: SUC P2 2015-2018. *La Revue électronique suisse de science de l'information (RESSI)*, 16, December 2015. http://www.ressi.ch/num16/article_110
- [2] **BURGI, P.-Y., BLUMER E., MAKHLOUF-SHABOU, B.** (2017). Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs. *International Federation of Library Associations and Institutions (IFLA) Journal*, 43, pp. 5-21. DOI 10.1177/0340035216678238
- [3] Par exemple, **CHAVAN V., PENEV L.** (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12(15): S2. doi:10.1186/1471-2105-12-S15-S2; voir aussi **TENOPIR C. et al.**, (2015) Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE* 10(8): e0134826. doi:10.1371/journal.pone.0134826.
- [4] <https://www.dlcm.ch>
- [5] Il s'agit en particulier des *Laboratory Information Management Systems (LIMS)* et *Electronic Laboratory Notebook (ELN)*.
- [6] **FÜRHOLZ, A., JAEKEL, M.** (2021) Data Life Cycle Management Pilot Projects and Implications for Research Data Management at Universities of Applied Sciences. *La Revue électronique suisse de science de l'information (RESSI)*. N° 22.



- [7] **BEZZI, M.** (2020). Préservation des données de recherche : proposer des services de soutien aux chercheurs du site Uni Arve de l'Université de Genève. Mémoire de Master : Haute école de gestion de Genève.
- [8] **WALLIS J. C., ROLANDO E. and BORGMAN C. L.** (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS ONE 8(7).
[DOI:10.1371/journal.pone.0067332](https://doi.org/10.1371/journal.pone.0067332).
- [9] ISO 14721:2012 (2012). Space data and information transfer systems – Open archival information system (OAIS) – Reference model
- [10] FAIR pour “Findable”, “Accessible”, “Interoperable”, et “Reusable”.
- [11] **BURGI, P.-Y, CAZEAUX, H., ECHERNIER, L.** (2019). A versatile solution for long-term preservation of research data: Data Life-Cycle Management: the Swiss Way. In: iPRES - 16th International Conference on Digital Preservation. Amsterdam (The Netherlands).
- [12] <https://datacite.org>
- [13] <https://creativecommons.org> - par défaut, nous recommandons CC-BY et CC0.
- [14] PREservation Metadata: Implementation Strategies (PREMIS)
<https://www.loc.gov/standards/premis>
- [15] Open Archives Initiative Protocol for Metadata Harvesting -
<https://www.openarchives.org/pmh>
- [16] L'environnement de développement Jupyter Notebook est très utilisé par les chercheurs car il permet de mélanger de la documentation avec du code de programmation - <https://jupyter.org>
- [17] Archiving Your Research Data with Yareta - IT Service Catalog - UNIGE -
<https://yareta.unige.ch>
- [18] Home - OLOS: The Swiss solution for managing research data - UNIGE -
<https://olos.swiss>

