



Bilan scientifique Dataacc mars 2019 – novembre 2020

Objectifs

Lancé au printemps 2019, le dispositif DATAACC' propose une approche originale d'accompagnement des chercheurs en s'inscrivant dans une logique double. D'une part, l'appui à la gestion des données dans un contexte de science ouverte. D'autre part, la volonté de construire un service taillé sur mesure pour les chercheurs issus des deux disciplines pour lesquelles l'UCBL et l'UGA disposent de fonds documentaires d'excellence : la chimie et la physique. La première phase de déploiement du projet a permis de tester cette approche hybride auprès des chercheurs, qui sont parties prenantes du dispositif.

Méthode

Les premiers mois du projet ont été consacrés à un important travail de repérage des pratiques des chercheurs des deux disciplines, afin de comprendre les types de données produites et les besoins qui pouvaient en découler. Cette prise de contact a permis de sensibiliser nos interlocuteurs aux objectifs du projet Dataacc afin de les associer aux différentes étapes. Les entretiens, répartis à parts égales entre la chimie et la physique, ont été réalisés auprès de 30 chercheurs de septembre 2019 à octobre 2020.

| Laboratoires concernés |
|--|
| Laboratoire des Matériaux et Interfaces (Université Lyon 1) |
| Institut Lumière Matière (Université Lyon 1) |
| Institut des Matériaux et Polymères (Université Lyon 1) |
| Institut de Chimie et Biochimie Moléculaires et Supramoléculaires (Lyon 1) |
| Labo de chimie de l'ENS Lyon |
| IRCELyon (Université Lyon 1) |
| Chimie Catalyse Polymères et Procédés (Université Lyon 1) |
| Institut d'Électronique et des Systèmes (Université de Montpellier) |
| Centre européen pour la recherche nucléaire |
| European Synchrotron Radiation Facility |
| Institut Neel (Université Grenoble Alpes) |
| Institut des sciences de la Terre (Université Grenoble Alpes) |

| |
|--|
| Laboratoire d'Electrochimie et de Physico-chimie des Matériaux et des Interfaces (Université Grenoble Alpes) |
| Laboratoire Génie des Procédés Papetiers (Université Grenoble Alpes) |
| Laboratoire des Matériaux et du Génie Physique (Université Grenoble Alpes) |
| Laboratoire National des Champs Magnétiques Intenses (Université Grenoble Alpes) |
| Institut de Microélectronique Electromagnétisme et Photonique (Université Grenoble Alpes) |
| Laboratoire de Physique et Modélisation des Milieux Condensés (Université Grenoble Alpes) |
| Laboratoire Science et Ingénierie des Matériaux et Procédés (Université Grenoble Alpes) |
| Laboratoire Modélisation et Exploration des Matériaux (Université Grenoble Alpes) |

Réalisations

I- Lancement de la plateforme Dataacc

Conception du site

Étape emblématique du projet, la [plateforme Dataacc](#) a officiellement été mise en ligne le 30 avril 2020, avec trois mois d'avance par rapport au calendrier initial, qui annonçait un lancement en juillet. La démarche de conception du site a été facilitée par l'agence Iris Interactive, basée au Puy-en-Velay et sélectionnée à l'issue d'une mise en concurrence d'une dizaine d'agences nous ayant transmis leur candidature (devis et cahiers techniques). Le choix du CMS s'est porté sur Wordpress, outil open source largement utilisé par les communautés.

Le site a été pensé autour de quatre grands besoins reprenant le cycle de vie des données et identifiés comme suit :

- Trouver des données
- Valoriser ses données
- Documenter ses données
- Conserver ses données

S'y ajoute une rubrique plus classique, "se former" qui permet de recenser des sites proposant des informations adaptées aux deux disciplines.

Ces clés d'entrée sont complétées par un menu plus détaillé déclinant plusieurs rubriques : Actualités, Bonnes pratiques, Besoins, Réseau et Informations générales sur le projet.

Des développements plus spécifiques ont été demandés à l'agence Iris afin de mettre à disposition des contenus répondant aux besoins des chercheurs sur l'identification d'entrepôts de données ou de logiciels de cahiers de laboratoire électroniques.

Une version en anglais de Datacc a par ailleurs vu le jour en février 2021, ce qui permettra de toucher les doctorants et chercheurs non-francophones sur la problématique de gestion des données.

Des ressources inédites

A travers la plateforme, les chercheurs en chimie et physique disposent désormais de ressources remettant en contexte les enjeux des plans de gestion de données, assorties d'informations pratiques facilitant la rédaction de ces derniers. Ils bénéficient également d'un [outil d'aide à la section d'entrepôts disciplinaires et généralistes](#) susceptibles d'accueillir leurs données. Un tel repérage disciplinaire en chimie et en physique, avec description précise des politiques des entrepôts de données, n'avait encore jamais été réalisé à ce jour. Outre les entrepôts, une classification des *data journals* est également proposée, contribuant ainsi à sensibiliser les chercheurs à un mode de publication émergent encore peu connu et pourtant envisagé comme un levier à actionner dans le cadre du plan national pour la science ouverte.

Outre le volet consacré à l'ouverture des données, le site de Datacc propose également une typologie très détaillée des logiciels de cahiers de laboratoire électroniques adaptés à la chimie voir à la physique pour l'un d'entre eux. Là encore, si des benchmarks de ce type de logiciels existent en ligne, le format proposé est inédit. Par ailleurs, les solutions LABY et SciFormation n'ont, à ce jour, été décrites dans aucun benchmark public en dehors de celui diffusé sur le site de Datacc. Ce benchmark a largement été réutilisé dans le cadre des travaux menés sur les cahiers de laboratoire électroniques à l'échelle du CNRS et du Comité pour la Science Ouverte (CoSO).

Un site co-construit avec les chercheurs

Autre point à souligner : plusieurs ressources ont été conçues en **collaboration étroite avec les chercheurs**. Il en va ainsi des deux infographies réalisées par DoraNum, pour lesquelles les scénarios ont été adressés à des chercheurs, physiciens et chimistes, invités à commenter et enrichir l'existant. [L'infographie sur les cahiers de laboratoire électroniques](#) n'aurait pas pu aboutir sous cette forme sans le concours précieux de Thierry Billard, enseignant-chercheur en chimie de synthèse à Lyon 1, qui nous a fourni un exemple d'expérience issu de son propre cahier de laboratoire. Enfin, la classification des formats de données recommandés en chimie ainsi que l'article consacré à la documentation des données via du code informatique ont respectivement bénéficié de l'appui de Thierry Billard et Colin Bousige, chercheur en physique au laboratoire des matériaux et des interfaces à Lyon 1.

Le déploiement du projet a par ailleurs été l'occasion de tisser des relations avec des doctorants particulièrement concernés par l'enjeu de la reproductibilité ou l'usage des cahiers de laboratoire. [Six vidéos](#) ont ainsi été tournées puis mises en ligne sur le site entre le printemps et l'automne 2020. Elles illustrent de manière percutante les problématiques auxquelles les jeunes chimistes et physiciens sont confrontés lorsqu'ils tentent de reprendre des travaux antérieurs pour nourrir leur propre thèse.

Des actualités et des événements qui dynamisent la plateforme

Soucieuse de tenir la plateforme à jour, l'équipe-projet produit chaque mois de nouveaux contenus venant alimenter les rubriques Actualités et Événements du site. Au fil des mois, le [bulletin de veille](#) est devenu un rendez-vous éditorial incontournable, lancé dès avril 2020, en

parallèle de la mise en ligne du site le 30 avril. La visibilité du bulletin s'est accrue considérablement à partir de septembre 2020, avec le lancement d'une [newsletter](#) qui compte à ce jour 254 inscrits.

La rubrique Événements du site, qui permet de relayer les webinaires, conférences ou formation en lien avec la gestion des données, tout en privilégiant une approche disciplinaire, est elle aussi constamment alimentée.

II- Des événements contribuant à fédérer la communauté

Outre la plateforme Dataacc, l'équipe-projet a organisé plusieurs manifestations en direction des chercheurs visant à jeter les bases d'un réseau de chercheurs-référents sur les différents enjeux abordés.

Un premier événement, baptisé [data lunch](#) et organisé sur le campus de la Doua à Villeurbanne, a ainsi attiré une trentaine de participants le 17 décembre 2019, venus assister à deux interventions de chercheurs sur les cahiers de laboratoire électroniques. Ce rendez-vous a permis de nouer des relations précieuses avec les chercheurs présents, qui se sont consolidées au fil du temps. Il a été suivi de quatre webinaires (15/1/2020, 31/03/2020, 24/04/2020 et 7/5/2020), dont le format, dicté par les conditions sanitaires à partir de mars 2020, a fourni une réponse pragmatique aux circonstances imposées.

Le lancement d'un accompagnement à la rédaction de plans de gestion de données à l'Université Grenoble Alpes a été précédé par une demi-journée d'échanges le 7 juillet 2020 autour des Plan de Gestion des données, organisée en distanciel avec l'INIST et l'Infrastructure de Calcul Intensif et de Données GRICAD. Les 32 participants ont pu, par petits groupes, examiner un plan de gestion des données et aborder de nombreux points problématiques pour les chercheurs

Afin faire écho du projet auprès de la communauté en physique, Bart van Tiggelen, directeur de recherche (laboratoire LPMMC) associé au projet depuis le début, a organisé, pour 3 des plus importants laboratoires grenoblois en physique (LPMMC, Institut Neel, et LNCMI), une rencontre autour de Dataacc le 18 septembre 2020. Cette demi-journée, qui a rassemblé une vingtaine de chercheurs, a ensuite déclenché l'accompagnement concret de deux projets autour de la gestion des données (description, diffusion).

III- Cahiers de laboratoire électroniques

Quatre solutions de cahiers de laboratoire électroniques ont été testées tout au long de l'année 2020. Il s'agit d'elabFTW, Chemotion, Mbook et LabCollector. Les deux premiers logiciels sont open source et les deux autres propriétaires. Une cinquantaine de chercheurs ont participé aux tests, principalement répartis entre les Universités de Lyon 1, Grenoble et Chambéry. Cette étape s'est soldée par un premier retour d'expérience de la part des utilisateurs, qui permet de faciliter l'engagement de la deuxième phase du projet sur 2020-2022.

L'ouverture d'instances de tests, couplée aux webinaires de démonstration organisés par Dataacc, a permis d'accompagner les chercheurs dans la découverte d'outils qu'ils ne connaissaient pas, tout en mesurant les différences notables et le temps de prise en main nécessaire de ces logiciels, en remplacement du traditionnel cahier papier.

IV- Accompagnement aux plans de gestion des données

Partie intégrante du service mis en place, l'accompagnement des chercheurs à la rédaction de plans de gestion (et d'ouverture) de données en physique et en chimie a amené l'équipe-

projet à intervenir auprès de 8 laboratoires entre 2019 et fin 2020. La liste est disponible ci-dessous :

| Laboratoire - projet | Exemples d'actions menées |
|---------------------------------|--|
| IRCELYon - CatReMo | Aide à la rédaction du PGD à partir de la plateforme DMPOpidor. Relecture et correction du PGD. |
| ISA - Lithogreen | Action de sensibilisation à la rédaction de PGD et l'utilisation de Zenodo. |
| C2P2 - Dynamex | Accompagnement sur les enjeux de stockage et d'archivage des données. |
| Cermep | Organisation d'un atelier sur les PGD et les entrepôts de données. |
| Laboratoire Pasteur - ENS Paris | Organisation d'un séminaire à destination des chercheurs du laboratoire sur les plans de gestion de données et les cahiers de laboratoire électroniques. |
| LNCMI | Aide à la rédaction du PGD d'un projet ANR et d'un projet européen |
| SIMAP | Présentation de Datacc et échange sur les pratiques de gestion des données et sur les résultats négatifs ou non-concluants |
| ISTerre | Présentation de Datacc' et des cahiers de laboratoire électroniques |

V- Bilan indicateurs

| | |
|---|---|
| Statistiques de fréquentation du site | 2 563 visiteurs - 20 718 pages vues |
| Appropriation de l'outil cahier de laboratoire | 31 testeurs d'elabFTW en 2020, 15 testeurs Mbook, 5 testeurs LabCollector, 5 testeurs Chemotion. |
| Nombre d'immersions et entretiens avec les chercheurs | 30 |
| Nombre de chercheurs participant aux événements | 25 (1er data lunch), 15 (pour chacun des quatre webinaires sur les ELN), 20 (séminaire ENS), 32 (demi-journée UGA sur les PGD), 22 (demi-journée UGA présentation Datacc'). |

VI - Communication

Des retours d'information réguliers ont été effectués auprès des services de communication de CollEx-Persée, afin d'alimenter la [page projet de Datacc](#), ainsi que le [rapport d'activité annuel](#).

L'équipe-projet a également participé à l'étape lyonnaise de l'Opidor Tour en janvier 2020, lequel a constitué un excellent moyen de faire connaître Datacc auprès des nombreux chercheurs présents dans l'amphi du CNRS.

Au plan institutionnel, plusieurs relais de communication ont été mobilisés. Les deux Universités ont ainsi co-signé un [communiqué de presse](#) publié le 25 mai 2020, pour annoncer le lancement de la plateforme Dataacc.org. Cette information a déclenché la publication d'une dépêche AEF le 29 mai 2020 intitulée : "Les universités Grenoble Alpes et Lyon-I lancent une plate-forme pour l'ouverture des données en physique et chimie".

A Grenoble, le site du Data Institute a par ailleurs publié un [focus sur Dataacc](#), revenant sur les buts poursuivis et la coopération menée avec les chercheurs.

La BU de Lyon 1 relaie quant à elle le projet sur le site à travers une [page dédiée](#).

Le projet Dataacc a en outre fait l'objet d'une présentation lors de l'édition de janvier 2020 des DIRBUIST, avec une intervention commune d'Isabelle Eleuche (directrice du SCD de Lyon 1) et Frédéric Saby (directeur de la BAPSO - UGA).

Enfin, l'équipe-projet relaie régulièrement les actualités du projet et de la plateforme sur les fils twitter maintenus par [Marie-Emilia Herbet](#) et [Lucie Albaret](#).

Quelques retours des usagers des services de Dataacc :

"Grand merci pour l'envoi, toujours aussi riche et très intéressant ! Vous maintenez la qualité de la lettre à un très haut niveau." Un chercheur en chimie de Lyon 1.

"C'est ce genre d'infographies qu'il faudrait montrer aux responsables qui auront, au final, le dernier mot sur les cahiers de laboratoire électroniques. C'est assez parlant." Un ingénieur d'études en chimie à Lyon 1.

"Encore merci pour votre aide pour la version 1 du plan de gestion de données, qui a été validée par l'ANR courant juillet." Un chercheur en chimie de Lyon 1.

VII- Enseignements tirés de la première phase de mise en oeuvre de Dataacc

Les observations et les entretiens réalisés ont permis de recueillir beaucoup d'éléments permettant d'ajuster les contenus de la plateforme et les contours des services proposés. Les chercheurs ont pour la plupart répondu de manière favorable, voire enthousiaste aux sollicitations de l'équipe-projet. Ils ont apprécié la démarche entreprise, la possibilité ainsi offerte de s'exprimer sur leur façon de gérer les données ainsi que la qualité des échanges sur cette thématique.

Il est à noter que, bien plus que le respect des principes de la science ouverte, c'est la possibilité de mettre en œuvre et de manière concrète une "bonne" gestion des données dans le cadre d'un projet, d'une équipe, d'un doctorat qui intéresse le plus les chercheurs, afin d'en faciliter la réutilisation. A plusieurs reprises, la question du partage des codes s'est posée, ces derniers pouvant être plus intéressants pour la communauté que les données brutes. Évidemment, la publication d'articles avec certaines données associées est la voie privilégiée par les chercheurs rencontrés.

Au-delà des constats dressés sur le type et l'usage des données produites en chimie et physique, la force du projet Dataacc réside dans sa capacité à revisiter complètement le rapport entretenu entre les bibliothèques et les chercheurs. Ces derniers se sont montrés agréablement surpris du positionnement "inattendu" de la bibliothèque sur les cahiers de laboratoire électroniques. De par leur nature, l'usage de ces derniers semblait exclusivement réservé aux chercheurs. Néanmoins, les bibliothèques ont su s'approprier rapidement cet enjeu, au point de devenir un animateur de communauté en la matière. Cette médiation s'est révélée efficace dans le sens où elle a permis de faire avancer la réflexion des laboratoires sur leurs pratiques, sans exposer les plus convaincus du bien-fondé des cahiers de laboratoire électroniques par rapport à d'autres chercheurs plus dubitatifs.

Outre le “pari” mené sur les cahiers de laboratoire électroniques, nos différentes interventions sur les plans de gestion de données permettent de mesurer l’étendue du chantier qui s’ouvre. De manière générale, nous avons pu constater la méconnaissance ou la perplexité face à l’exercice demandé. Plutôt qu’un outil scientifique au service de la traçabilité, les plans de gestion de données restent souvent perçus comme un document administratif venant alourdir la charge bureaucratique pesant sur les chercheurs.

Nos accompagnements ont donc consisté à aborder le sujet de manière pragmatique dans le but de faire gagner du temps aux chercheurs. Ces échanges nous ont également permis de diffuser des informations souvent méconnues sur l’existence d’outils permettant de déposer et/ou de trouver des données de recherche.

Plus largement, ces accompagnements “personnalisés” et la plateforme Datacc’ ont permis de légitimer les bibliothèques dans ce rôle de soutien à la gestion des données de la recherche.

Le projet Datacc a également permis de nouer des collaborations avec les acteurs impliqués dans la gestion des données. Nous avons ainsi été invités à participer à deux groupes de travail du Comité pour la Science Ouverte dont les lettres de mission sont axées sur l’appropriation de la science ouverte par les communautés et le déploiement des cahiers de laboratoire électroniques. Nous participons ainsi de manière active aux livrables, en capitalisant sur l’expérience acquise dans le cadre de Datacc.

Des rapprochements ont également eu lieu avec le réseau des centres de documentation des instituts IN2P3 (réseau Démocrite) et INSU sur la démarche engagée. Plusieurs retours ou suggestions de collègues ont permis ou permettront d’améliorer la qualité des informations sur le site. La démarche de Datacc (proposer des contenus adaptés à deux communautés) a également intéressé l’UMS GRICAD, Infrastructure de Calcul Intensif et de Données de Grenoble. Cette UMS est à l’initiative de la constitution de la cellule data Grenoble Alpes, qui se propose d’accompagner les chercheurs sur tous les aspects de la gestion des données. S’appuyant sur l’expérience acquise dans le cadre de Datacc, la BAPSO de Grenoble participe donc activement aux travaux de cette cellule. Côté Lyon 1, où la question des données de recherche est structurée différemment, le projet Datacc a permis de nouer une relation de travail quotidienne avec la référente science ouverte de l’établissement Cherifa Boukacem, depuis sa désignation à ce poste à l’automne 2020.

Conclusion

La première phase de mise en œuvre du projet Datacc a permis de resserrer les liens entre les bibliothèques et les communautés de chercheurs en chimie et physique, dans un contexte où les services de documentation s’interrogent constamment sur leur rôle à l’égard de ces publics. Le dernier rapport Couperin sur les pratiques documentaires des chercheurs est un puissant révélateur du risque encouru à ne pas déployer des services adaptés aux besoins et aux pratiques des chercheurs (portail documentaire sous-utilisé, discrédit potentiel du rôle du bibliothécaire etc.)

A travers une approche non-dogmatique, fondée sur les outils (plateforme Datacc, cahiers de laboratoire électroniques, newsletters) et la proximité (rendez-vous, événements en ligne ou en présentiel), l’équipe-projet a pu, à son échelle, créer une relation de confiance avec les chercheurs concernés.

La deuxième phase de déploiement du projet, facilitée par l’appui renouvelé du GIS-CollEx Persée, sera l’occasion de conforter les axes mis en œuvre, tout en développant un nouveau champ de coopération fondé sur les résultats scientifiques dits “négatifs”.