FOSTER *Facilitate Open Science Training for European Research*
« Open Access et gestion des données dans Horizon 2020 »

Synthèse des journées tenues à Paris les lundi 29 et mardi 30 juin 2015

CARTIER, Aurore, Université Paris Descartes
MOYSAN, Magalie, Université Paris Diderot
REYMONET, Nathalie, Université Paris Diderot

Deux journées FOSTER dédiées à l'Open Access et à la gestion des données de la recherche dans les projets Horizon 2020 se sont déroulées à l'Université Paris Diderot les lundi 29 et mardi 30 juin.

Porté par le Service commun de la documentation (SCD) de l'Université Paris Descartes en collaboration avec la Direction d'appui à la recherche et l'innovation (DARI) et le Bureau des archives de Paris Diderot, ces deux journées répondaient aux appels à projet du programme FOSTER *Facilitate Open Science Training for European Research*. FOSTER est lui-même financé par la Commission européenne et porté par un consortium européen dans l'objectif d'accompagner et d'aider au financement d'actions de formation en faveur du libre accès et de la gestion des données dans les appels du programme cadre Horizon 2020.

L'engagement européen en faveur du libre accès et des démarches d'« Open Science » n'est pas nouveau mais se renforce d'année en année en passant de logiques d'encouragement à celles d'obligation. Les récentes déclarations du gouvernement français concernant sa stratégie numérique le 18 juin dernier vont également dans ce sens puisque « La France accentue son engagement dans l'ouverture des publications et des données de la recherche financée sur fonds publics[1] ». À cet égard, le Comité d'éthique du CNRS (COMETS) souligne également la nécessité, aujourd'hui, de former « Les chercheurs et les personnels du monde de la recherche aux dimensions éthique de la gestion des données, en particuliers au respect de la vie privée, de la vie propriété intellectuelle, de la qualité et de l'intégralité des données[2] ».

C'est donc dans ce contexte que s'inscrit la démarche de formation mise en place au travers de ces deux journées. La première journée devait permettre d'appréhender des concepts ainsi que les exigences de la Commission européenne en matière de libre accès, la seconde était dédiée aux différentes méthodes et outils pour s'y conformer.

Le public visé était donc les chercheurs, bien sûr, mais aussi les ingénieurs-projets et les professionnels de l'IST qui les accompagnent dans leurs travaux.

[1] République Française, Premier Ministre. *Stratégie numérique du gouvernement*, p.15 (juin 2015)
[2] COMETS. *Les enjeux éthiques du partage des données scientifiques* (7 mai 2015)

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Les échanges nés de ces rencontres ont ouvert la voie à de nouvelles coopérations et ont permis de poser un socle de compétences désormais requises pour affirmer la compétitivité de nos établissements dans la sphère européenne, compétences parmi lesquelles figurent en bonne part *l'open access* et la gestion des données dont les professionnels de l'IST sont des acteurs historiques.

En faisant volontairement appel à différents interlocuteurs issus d'institutions et de domaines professionnels variés (chargé de politique auprès de la Commission européenne, ingénieur projet, bibliothécaire, ingénieur de recherche, juriste, archiviste, etc.), l'objectif était de pouvoir couvrir l'ensemble des compétences nécessaires au montage d'un projet de recherche européen et d'identifier les acteurs susceptibles d'intervenir sur ces questions, de la candidature jusqu'à l'évaluation finale d'un projet.

À travers ces deux journées d'ouverture nationale, nous espérons avoir consolidé la performance de nos établissements dans les appels à projet européens, œuvré en faveur d'une science plus ouverte et convaincu de l'intérêt d'une nouvelle synergie entre les acteurs de nos établissements au service de la recherche.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

INTRODUCTION

Le Président de l'Université Paris Descartes, Frédéric Dardel et le vice-président du CA de l'Université Paris Diderot, François Villa, ainsi que Aurore Cartier, porteuse du projet pour le SCD Paris Descartes, ont introduit ces deux journées de formation.

Pour F. Villa, tout article est conçu pour être lu et discuté. Les publications servent aussi à évaluer la pertinence des financements octroyées et la carrière des chercheurs. Les publications permettent également l'information des professionnels et du public en dehors des bibliothèques non-spécialisées.

Pour F. Dardel, la construction de l'Espace européen de la recherche passe aussi par la publication et la production de données scientifiques. Une révolution de la publication a eu lieu avec le numérique, mais la science également a changé. La production des données est devenue un sujet majeur dans toutes les disciplines : autrefois la production de données était coûteuse, aujourd'hui stocker de la donnée est plus cher que la produire. On assiste ainsi à une inversion de la tendance qui incite à jeter les données, ce qui induit une défiance sur la fiabilité de la science si on ne peut pas vérifier les résultats. Financièrement, la donnée a une valeur, pourtant elle est un bien public car produite avec des financements publics. Pour toutes ces raisons, les scientifiques ont besoin d'accompagnement.

A. Cartier rappelle que ce projet de formation est né d'une double rencontre de communautés qui, bien qu'œuvrant au sein d'un même établissement ou d'une même communauté d'établissements, n'ont pas toujours connaissance de leur existence et expertise respectives : d'une part entre professionnels de l'information scientifique et technique (IST) issus d'horizons différents (bibliothécaires, archivistes, documentalistes) d'autres part entre professionnels de l'IST et experts en ingénierie de projets implantés au sein des directions de la recherche (ingénieurs-projets, juristes, chargés de valorisation). Impulser une synergie entre ces acteurs contribuera à former les personnels à la gestion des données.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Module 1

Comprendre les exigences de Horizon 2020 : *open access* des publications, ouverture et données de gestion

Christine OLLENDORFF (Directrice de la Documentation et de la Prospective, Arts et Métiers) : *Open Access : définitions, problématiques, panorama en France*

Christine Ollendorf est actuellement directrice de la Documentation et de la Prospective à l'ENSAM et animatrice du groupe de travail sur l'Open Access (GTAO) du consortium Couperin.

L'édition scientifique est en pleine structuration avec les groupes multinationaux en position de plus en plus oligopolistique, l'augmentation des frais de publication et les coûts exorbitants des abonnements dans le cadre des *big deals*.

L'open access (OA) peut être défini comme un accès libre, immédiat, permanent et en ligne aux articles scientifiques revus par les pairs. Mais ouvert n'est pas libre, les contenus peuvent être soumis à des droits. La conférence de Budapest (2002) avait défini les deux voies de l'open access :

- Voie verte : auto-archivage par les chercheurs ou archivage par une tierce personne des publications scientifiques dans des archives ouvertes. Selon le site Sherpa/Romeo, 76% des éditeurs autorisent maintenant le dépôt des publications en archive ouverte ;

- Voie dorée : publication d'articles dans des revues en libre accès, quel que soit leur mode de financement. Il peut s'agir d'un modèle auteur-payeur ou sans financement par l'auteur ou d'un modèle hybride entre les deux. Selon le site DOAJ, 64% des articles nécessitent des APC (*articles processing charges*, ou frais de publication), ce qui induit des inégalités entre chercheurs et incite les « éditeurs prédateurs » à escroquer les chercheurs. Quant au modèle hybride, il induit de payer deux fois (*double dipping*) : une fois l'abonnement à la revue et la deuxième l'APC pour chaque article publié.

Les archives ouvertes existent le début des années 1990. En 2013, la CE déclare un objectif de 60% de publications en open access en 2016 et 100% en 2020 car l'open access accélère le processus de recherche et d'innovation, mais aussi favorise le taux de citation, favorise la prise en main de la science par les profanes. Mais il y a peu de dépôts volontaires de la part des chercheurs. Le chercheur souhaite conserver sa souveraineté scientifique et publier où et quand il veut, il est intéressé par la notion de facteur d'impact, mais il est soumis à des contraintes qui forment un environnement complexe et difficile à appréhender : les chercheurs déposent peu parce qu'ils manquent d'information sur l'open access, sur les éditeurs, parce qu'ils manquent de temps, qu'il n'y a pas de politique nationale claire, etc. À ce sujet, le Rapport du Conseil National Numérique recommande (p.280) de « Faire de la publication ouverte une obligation légale pour la recherche bénéficiant de fonds publics » [3].

[3] http://www.cnnumerique.fr/wp-content/uploads/2015/04/2306_Rapport-CNNum-Ambition-numerique_sircom_print.pdf

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Jean-François DECHAMP (Commission européenne, Direction-Générale Recherche et Innovation) :
Open Access et gestion des données : quel agenda pour la recherche scientifique européenne ?

*Depuis 2005, Jean-François Dechamp est responsable de politiques au sein de la Direction-générale Recherche et Innovation de la Commission européenne à Bruxelles. Il s'occupe des questions liées au libre accès (*Open access*) et plus récemment au copyright et la fouille des données (*Text and Data Mining*). Il a travaillé auparavant pour différents groupes d'intérêts et organisations non-gouvernementales, en particulier en Italie et en Chine. Il a obtenu son Diplôme d'État de Docteur en Pharmacie auprès de l'Université de Strasbourg.*

La Commission européenne (CE) propose la législation et invite les États membres à légiférer, elle est aussi une agence de financement qui définit les règles de distribution et de participation avec une nécessaire cohérence avec le législatif, elle finance des projets pérennes et des infrastructures.

Pour la CE, l'open access n'est pas une obligation de publier, ne va pas à l'encontre des brevets, et doit garantir la même qualité scientifique des publications. Les objectifs de la CE sont en effet d'optimiser l'impact des résultats de la recherche financée par fonds publics (efficience de la science), en mettant en place l'open access dans les projets financés, en finançant des projets comme Foster, en incitant à la collaboration et la coordination des États membres.

Les documents fondamentaux de juillet 2012, deux Communications ("*A reinforced ERA partnership for excellence and growth*" et "*Towards better access to scientific information: boosting the benefits of public investments in research*") et une Recommandation aux États membres sur l'accès et la conservation des documents scientifiques constituent les références de la stratégie de la CE pour l'open access. Deux *Guidelines*, publiés en décembre 2013, donnent les informations sur la manière de procéder pour l'open access des publications (obligatoire, avec un embargo de 6 ou 12 mois) et la gestion des données dans Horizon 2020 (soutenue, mais non obligatoire). Voici ce qui est devenu une obligation contractuelle dans H2020, lorsqu'il y a publication de résultats :

- déposer une copie électronique de la publication (même si elle est initialement en *gold* OA), de façon à être moissonnable par OpenAIRE, l'archive ouverte de la CE ;
- s'assurer qu'elle soit accessible dans le délai de 6 ou 12 mois (selon les disciplines) ;
- rendre les métadonnées descriptives accessibles, par le biais de l'archive ouverte choisie ;
- « tendre au dépôt » des données qui sous-tendent les publications (*underlying data*), même si ces données ne sont pas en open access.

Concernant les données de la recherche, le pilote *Open Research Data* porte sur 7 domaines thématiques, mais n'importe quel projet peut aussi choisir de venir volontairement en *opt-in* ou en sortir (*opt-out*) à n'importe quel moment du projet, par amendement : le pilote est flexible. La participation au pilote ne compte pas dans l'évaluation.

Le participant va déposer dans le réservoir de son choix, doit « prendre des mesures » pour exploiter les données. Un plan de gestion de données, ou *Data management plan* (DMP) est obligatoire pour les participants au pilote et optionnel pour les autres. Ce DMP peut être bref et décrit :

- comment les participants vont gérer les données : concerne tous les projets, même hors du pilote ORD ;

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

- la stratégie pour la gestion de la connaissance et sa protection :
    - prévoir les publications en open access dès qu'il y a publication (art. 29 de la convention de financement) ;
    - exploiter les résultats, pas forcément de façon commerciale : une fois protégés, ils doivent être diffusés (art.28) ;
- pilote ou *opt-in* : obligation de publier les données ou s'en expliquer dans le DMP (29.3).

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Module 2

Comprendre les exigences d'Horizon 2020 : propriété intellectuelle et industrielle

Nathalie LE BA (DAJ CNRS) : *La propriété intellectuelle dans les projets collaboratifs européens*

Juriste, spécialisée en Affaires européennes et Propriété intellectuelle Nathalie Le Ba travaille actuellement au sein du Pôle Accords Propriété Intellectuelle de la Direction des Affaires Juridiques du CNRS.

La propriété intellectuelle regroupe la Propriété littéraire et artistique, dont le Droit d'auteur, les Droits voisins et le Droit *sui generis* sur les bases de données, et la Propriété industrielle.

La publication est concernée par le droit d'auteur dès sa création. Les chercheurs n'ont pas besoin de demander l'autorisation de leur administration pour divulguer leur œuvre par publication et ils sont titulaires de leurs droits moraux et d'auteur. Cependant, avant de publier il peut être opportun de s'interroger sur le droit d'auteur (contrefaçon d'une autre œuvre) et le droit moral (par la citation des co-auteurs et auteurs), mais aussi sur les autres intérêts en jeu : préserver les résultats ou le savoir-faire mis en œuvre qui pourraient donner lieu à un dépôt ultérieur (antériorité) ; engagement auprès de tiers comme par exemple dans les accords de consortiums de projets européens (auquel cas il pourrait être souhaitable de retarder la publication) ; obligations prévues par la convention de subvention comme par exemple la mention du numéro de la convention et la modalité de publication en open access.

Points à respecter avant de réaliser un dépôt en archives ouvertes :

- s'assurer de l'accord des autres co-auteurs et/ou contributeurs ;
- s'assurer de la non-confidentialité de la publication ;
- s'assurer que le document n'a pas déjà été confié à un éditeur ou un diffuseur avec un contrat spécifiant une non-divulgation :
    - s'il n'existe pas de contrat d'édition, il n'y a pas d'obstacle au dépôt en archive ouverte
    - si un contrat d'édition a été signé, il peut y avoir des limites :
        - s'il n'y a pas de cession exclusive des droits en particulier pour les supports électroniques, il n'y a pas d'obstacle au dépôt en archive ouverte
        - il peut être nécessaire d'attendre un certain délai avant de déposer (embargo) ce qui oblige ainsi à faire la balance entre : l'obligation de la CE de faire de l'open access après un embargo de 6 ou 12 mois, et ce que l'éditeur exige et qui peut être différent ;
        - en cas de cession exclusive des droits, il reste possible de solliciter l'éditeur pour lui demander son autorisation pour le dépôt.
- consulter les sites comme Sherpa/Romeo ou Héloïse pour connaître les exigences des éditeurs ;
- négocier, par avenant si nécessaire, la cession de droit non exclusive, y compris pour les illustrations au sein des publications.

Données de recherche : la propriété des données est déterminée dans l'accord de consortium, d'où l'importance de le négocier bien en amont du projet : cet accord pourra définir quel jeu de

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

données sera ouvert et quel autre sera protégé pour des raisons de confidentialité, de données à caractère personnel, de priorité d'exploitation, etc.

Les bases de données peuvent être protégées par le droit d'auteur si elles sont originales (sur le contenu) ou avec une architecture originale (appréciation au cas par cas). La possibilité d'accès à leur contenu ne donne pas nécessairement le droit de l'utiliser.

Nicolas GIRARDIN (SATT idfinnov) : *Les droits de propriété industrielle*

Nicolas Giradin est responsable propriété intellectuelle à la SATT Ile-de-France Innov. Auparavant, il a été ingénieur-brevet et consultant en bio-technologies.

La Propriété industrielle comprend les Droits sur les créations nouvelles, dont les brevets.

Le brevet est un titre de propriété ayant une durée de validité de 20 ans, sous réserve de paiement de taxes de délivrance ou d'annuité. Il fonctionne avec une protection territoriale, par pays ou par zones géographiques. Il revient à un droit d'interdire à un tiers d'exploiter l'invention telle que décrite. Il couvre une invention, c'est-à-dire une solution technique à un problème technique. Le brevet est un document descriptif de l'invention, de l'objet de la protection demandée (revendications), des caractéristiques techniques de l'invention, et peut comporter des illustrations.

Un logiciel porte l'originalité de son auteur dans son code source, mais il décrit aussi une méthodologie qui peut, elle, être protégée par un brevet.

Publication et brevet ne sont pas incompatibles, il reste possible de protéger une invention par demande de brevet avant que la publication ne soit diffusée, puis de publier ou disséminer, comme le requiert H2020. Le dépôt d'une *demande de brevet* permet ainsi de publier bien avant la délivrance du brevet.

Les droits au brevet : l'inventeur est la/les personne(s) physique(s) concernées, le titulaire (déposant) est le détenteur des droits, le propriétaire. Le plus souvent, le titulaire est l'employeur lorsque l'inventeur est salarié. Celui-ci doit déclarer son invention à son employeur avant toute publication ou divulgation à des tiers.

Les brevets dans les projets H2020 :

- la dissémination est souhaitable sous réserve de protection ;
- l'accord de consortium entre les partenaires au projet décrit ce que chaque partenaire possède comme connaissances en propre et/ou connaissances en copropriété ;
- exploitation : de même, l'accord de consortium décrit les résultats en propre et les résultats communs ainsi que leur modalité d'exploitation (exclusivité).

Le coût du brevet est éligible pendant la durée du projet, tout comme le coût de la publication.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Mariama COTTRANT (Université Paris 13) : *Le plan de diffusion et de valorisation des résultats et la notion d' "impact" dans les projets H2020*

Diplômée en Études européennes de King's College London et UCL, Mariama Cottrant est en charge du montage et du suivi des projets européens au SAIC de l'Université Paris 13 depuis 2013. Elle fait également partie du consortium du Point de Contact National « Technologies Futures et Emergentes » (FET).

Dans H2020, la notion d'impact est liée à la Stratégie de croissance de l'Union européenne : « L'Union doit devenir une économie *intelligente*, *durable* et *inclusive*… ». La notion d'impact est également liée aux constats faits à la fin du FP7 : « … l'Europe doit (...) augmenter les interactions entre les actions de recherche et d'innovation et les politiques liées à ces défis ». Les constats portent aussi sur la nécessité de renforcer sa recherche fondamentale et exploratoire ainsi que ses infrastructures, avec pour objectif de faire sortir la recherche des laboratoires, pour un impact sur la recherche, l'économie et la société. Le financement de H2020 doit donc être efficace en termes d'impact sur la recherche (excellence, mobilité et formation des chercheurs), impact économique (compétitivité européenne, création d'emplois), impact sociétal (politiques publiques, société civile).

Le lien avec l'open access est que l'information étant déjà financée par le contribuable au moment du financement de la recherche, ce contribuable ne devrait pas avoir à payer encore pour y accéder et l'utiliser.

L'architecture de H2020 en trois piliers (excellence scientifique, primauté industrielle et défis sociétaux) induit les attendus d'un projet retenu en termes d'impacts : quel est l'objectif général du pilier dans lequel se situe l'AAP ? quel type d'actions est ainsi financé ? quels sont les publics-cibles des résultats, l'impact à court et long-terme du projet, ainsi que le niveau de maturité (p/r à l'industrialisation des résultats) ?

Dans l'évaluation d'un projet H2020, l'impact compte de manière décisive (30% environ). Il est pertinent d'en faire un *work package* et de l'inclure dans l'ensemble du projet, pour tenir compte de la description de la dissémination et de sa mise en œuvre :

- dissémination : diffusion des *résultats* scientifiques ;
- communication : ne nécessite pas de *résultats*, on peut communiquer sur le *projet.*

Le *draft plan* de dissémination devra être mis en œuvre si le projet est financé : il devra décrire de façon claire, précise, concrète les modalités de diffusion, de gestion de la propriété intellectuelle : par exemple inclure un partenaire *end-user* (une entreprise qui prendra en charge l'exploitation des résultats).

Les obligations des projets financés décrites dans le Model Grant Agreement sont :

- Art. 27 : *Obligation to protect the results* s'il existe un potentiel d'exploitation industrielle ou commerciale ;
- Art. 28 : *Obligation to exploit the results* au moins dans d'autres recherches, ou par licence ;
- Art. : *Dissemination of results – Open Access*

Selon les objectifs et l'impact attendu du projet proposé, le porteur de projet va définir sa stratégie par rapport aux résultats attendus : Confidentialité/protection/exploitation, ou Communication/publication/diffusion. Ces deux démarches ne sont pas opposées mais l'option Communication peut tout à fait succéder à la phase d'exploitation voire la compléter, dans un timing différent. Du choix de la logique adoptée dépendra la teneur du plan de diffusion.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Module 3

Répondre aux exigences de Horizon 2020 : entrepôts, outils et solutions techniques de dépôt

Stéphane POUYLLAU (CNRS) : *Huma-Num, la TGIR des humanités numériques*

*Ingénieur de recherche au CNRS (Centre National de la Recherche Scientifique), Stéphane Pouyllau est spécialisé depuis 1999 en humanités numériques (*digital humanities*), en information scientifique et technique et en informatisation des données de la recherche en sciences humaines et sociales. Il est actuellement directeur-adjoint technique d'Huma-Num, la très grande infrastructure de recherche pour les humanités numériques.*

Les humanités numériques consistent à faciliter le tournant numérique de la recherche en SHS lors de la production et de la réutilisation de données numériques, à l'aide d'outils et de services adaptés.

La très grande infrastructure de recherche (TGIR) Huma-Num résulte d'un consortium entre disciplines et métiers, et a pour mission de définir des méthodes et des procédures et standards numériques partagés : Huma-Num permet de stocker, traiter et diffuser des données, mais aussi de les archiver à long-terme en coordination avec le CINES.

Elle regroupe un ensemble d'infrastructure et de systèmes informatiques mis à la disposition des équipes de recherche afin de mutualiser, diffuser et stabiliser dans le temps les données et documents issus de la recherche ainsi que les méthodes de traitement. L'attribution d'un identifiant pérenne permet de citer les données en les « exposant » au moissonnage d'autres plateformes par l'interopérabilité.

ISIDORE, *« Intégration de services, interconnexion de données de la recherche et de l'enseignement »*, est un service de collecte, enrichissement et d'accès aux documents et données numériques des sciences humaines et sociales, par moissonnage des notices, métadonnées et texte intégral de la production scientifique électronique, suivi d'un enrichissement.

NAKALA propose des services d'accès aux données par identifiant unique, et des services de présentation des métadonnées par exposition RDF.

Christine BERTHAUD (CCSD-CNRS) : *HAL, archives ouvertes*

Christine Berthaud est ingénieur de recherche et directrice depuis 2011 du Centre pour la Communication scientifique directe (CNRS). Elle représente la France au niveau européen sur les questions d'Open Access.

Le CCSD a pour mission de développer des archives ouvertes pour la communauté de l'ESR.

Une archive ouverte est un réservoir de données dans lequel on dépose des contenus scientifiques afin de les rendre librement accessibles sur le web ; l'auto-archivage consiste à déposer soi-même.

La convention de 2013 fait de Hal une archive inter-établissements (AMUE, CPU, organismes de recherche). Ces établissements déposent directement dans Hal ou s'engagent à reverser le contenu de leur archive institutionnelle dans Hal, qui assure l'archivage pérenne des contenus.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Hal est ainsi une plateforme commune pour garantir l'accessibilité aux contenus (le texte intégral), garantir un niveau scientifique homogène entre auteurs, donner de la visibilité internationale par la connexion formalisée avec d'autres plateformes, garantir l'archivage (auprès du CINES) à long-terme, au sein d'un réseau de partenaires et dans un lieu très sécurisé. Hal peut recevoir tout ce que la science peut produire comme types de document, ainsi que les données de type image, son, etc. Une vérification technique de contrôle qualité sur le document et sa cohérence avec les métadonnées est faite par l'équipe, ainsi qu'une vérification que le document est une production scientifique mais sans évaluation de contenu.

Dans le contexte d'Horizon 2020, Hal permet :

- la collecte du document publié ;
- la caractérisation du dépôt par des métadonnées descriptives, ces métadonnées permettront d'exploiter par la suite les documents de toutes sortes de façons.

Les métadonnées sont contrôlées par des référentiels de Hal, AuréHal, accessibles à tous avec un simple compte. Ces référentiels existent depuis une quinzaine d'années et sont corrigés et mis à jour quotidiennement ; ils constituent, de fait, des référentiels riches. Les projets européens sont référencés, on peut créer une collection dans Hal autour d'un projet européen. Hal est moissonné automatiquement par OpenAIRE, il est, de fait, inutile pour un chercheur français d'utiliser OpenAIRE.

Une construction est faite sur les URL pour disposer d'une URL simplifiée, qui permet d'accéder en priorité à la version la plus récente du document, et garantir la stabilité des URL dans le temps, quel que soit le lieu physique de stockage des documents.

L'identifiant unique des auteurs permet de regrouper les différentes variantes de nom d'auteur (Marie-Pierre Dupont, M.-P. Dupont, MP Dupont) et d'indiquer des identifiants issus d'autres applications telles que ResearcherID ou Tweeter afin de regrouper les identifiants numériques d'une même personne.

Dans Hal et à partir de son IdHal, un chercheur peut créer son CV qui présente ses activités, son parcours académique, afficher ses publications (toutes ou partiellement), les classer selon son choix, ajouter des widgets extérieurs ou mettre en avant ses projets de recherche, ses collaborations, les revues dans lesquels il publie, faire évoluer son CV en fonction de sa stratégie personnelle d'affichage, déposer ses publications dans Hal et référencer les publications dans ResearchGate : le CV de Hal est un outil de communication scientifique qui s'inscrit dans la pérennité et l'international. Il permet de renforcer l'identité numérique des chercheurs, mais aussi des laboratoires et des établissements.

Hal est une AO « ouverte » : très peu de documents sont sous embargo, les référentiels sont accessibles, il est possible de déposer des *preprints* pour la discussion entre chercheurs.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Natacha LECLERCQ-VARLAN (SCD, Université Paris Diderot) : *Publier en Open Access, quelles stratégies mettre en œuvre ?*

Conservatrice des bibliothèques, Natacha Leclercq-Varlan a rejoint le Service commun de la documentation de l'Université Paris Diderot-Paris 7 en 2011. Responsable de la documentation électronique pour l'ensemble des bibliothèques de l'Université, elle participe à des formations à destination des étudiants et des enseignants-chercheurs afin de faire connaître et promouvoir l'Open Access mais également pour les sensibiliser à la question du coût des ressources électroniques et les inciter à déposer leurs publications dans les archives ouvertes

Dans le cadre de H2020, publier en open access est une obligation, avec une incitation forte de publier aussi en open access les autres communications scientifiques (publications sans comité de lecture et autres supports de la communication scientifique : chapitre, livre, actes, poster, etc.) afin d'assurer la dissémination sur la plus large échelle possible et la pérennité des résultats scientifiques.

Où diffuser : dans une revue scientifique en open access (*gold oa*) ou dans une revue traditionnelle en déposant ensuite dans une archive ouverte (*green oa*). Ces modalités complémentaires répondent toutes deux à l'exigence de diffusion en open access de H2020. Même dans le cas d'une publication en *gold oa* le dépôt dès que possible du fichier dans un entrepôt d'archive ouverte reste une demande.

Le *gold* permet à l'auteur de bénéficier d'une validation par les pairs mais représente un coût ainsi qu'un risque financier avec les éditeurs prédateurs (escroquerie à l'édition) ; le *green* donne de la visibilité et la pérennité de l'accès, présente un coût direct nul, est souvent soumis à un embargo, la version à déposer n'est pas toujours le PDF de l'éditeur, mais les archives ouvertes sont interopérables et moissonnées par les moteurs de recherche. Pour sélectionner une revue dans laquelle publier en open access, utiliser les outils disponibles : DOAJ (répertoire de revues *gold*), liste des éditeurs prédateurs, site Sherpa/Romeo qui présente la politique des éditeurs en matière de *gold* et/ou de *green*, etc.

Déposer ses articles dans HAL permet d'une part d'alimenter OpenAIRE et d'autre part de créer une source pour générer sa liste de publications pour le *Final report* auprès de la CE à l'issue du projet.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

Module 4

Répondre aux exigences de Horizon 2020 : gestion, documentation et archivage

Aurore CARTIER (SCD, Université Paris Descartes), Magalie MOYSAN (Bureau des archives, Université Paris Diderot), Nathalie REYMONET (DARI, Université Paris Diderot) : *Établir un plan de gestion des données dans le cadre d'un projet européen*

Archiviste-paléographe (École nationale des chartes) et conservateur des bibliothèques (Enssib) spécialisée dans le domaine de l'Open Access et de l'Open Data, Aurore Cartier occupe actuellement des fonctions de responsable des services à la recherche et de coordinatrice des bibliothèques universitaires de médecine de l'université au sein du Service commun de la documentation de l'Université Paris Descartes.

Archiviste de formation (Université de Versailles-Saint-Quentin-en-Yvelines, promotion 2009), Magalie Moysan a travaillé dans diverses institutions (ministère, commune, entreprise), avant de rejoindre le Bureau des archives de Paris Diderot. Elle s'occupe de la collecte, du classement et de la valorisation des archives des composantes et des enseignants-chercheurs de l'établissement.

Documentaliste (INTD, 1992), Nathalie Reymonet travaille à la Direction d'Appui à la Recherche et à l'Innovation (DARI) de l'université Paris Diderot, sur l'open access et les indicateurs d'activité scientifique. Auparavant, elle a travaillé à la bibliothèque de l'Observatoire de Paris en tant que responsable de la section de Meudon, ainsi que chef de projet sur le catalogue de la bibliothèque et responsable des ressources électroniques.

Différents types de données sont produites ou collectées, avec des spécificités différentes. Aujourd'hui, lorsque des données primaires sont produites ou collectées au cours de la recherche, une part de celles-ci sont transformées et analysées, une faible part de ces dernières sont retenues pour l'exploitation et la production des résultats et une plus faible part encore de ces dernières sont publiées. Il y a donc une importante perte de données au cours du processus de publication.

Pourtant ces données dans leur ensemble présentent un intérêt scientifique si elles étaient réutilisées comme source ou pour fiabiliser des résultats, un intérêt financier car chères à produire, ainsi qu'un intérêt en termes de valorisation des travaux de l'établissement qui les a produites.

Pour ces raisons, la CE dans H2020 considère que les données doivent être détectables, accessibles, fiables, intelligibles, interopérables et réutilisables. Pour cela il importe de les décrire et les documenter, à l'aide d'un plan de gestion de données, ou *Data management plan* (DMP). Il s'agit d'un livrable pour la CE, décrivant la manière dont seront produites, traitées, décrites, diffusées et conservées les données au cours et à l'issue du projet. Dans H2020, le pilote *Open Research Data* prévoit que les bénéficiaires de financements produisent un plan de gestion, déposent les données décrites dans un entrepôt et les documentent. Les principaux champs d'un DMP sont :

- Responsabilité des données : répartition des rôles entre les partenaires du projet
- Ressources nécessaires à la mise en œuvre du DMP, y compris financières
- Description du/des jeu(x) de données
- Stockage, accès et sécurité des données – au cours du projet
- Métadonnées : documentation et organisation des données
- Dissémination des jeux de données – à l'issue du projet
- Sélection et archivage, que les données soient diffusées ou non

Il existe des outils pour détecter l'entrepôt le plus approprié pour les données spécifiques d'un projet. Déposer des données ne signifie pas forcément les ouvrir, il peut exister des exceptions

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

à leur diffusion (exploitation industrielle, confidentialité, secret défense, etc.) mais pas à leur description dans un DMP. Il existe des outils permettant de générer un DMP en ligne, ou d'être guidé dans la rédaction de ce livrable. Les compétences nécessaires à la rédaction d'un DMP sont multiples : chercheur qui produit les données, juristes qui rédigent les conventions de partenariat, professionnels de l'IST pour connaître les outils de mise en ligne et anticiper l'archivage à long terme des données.

Lorène BÉCHARD (CINES) : *La conservation pérenne des données de recherche*

Lorène Béchard est archiviste et responsable fonctionnel au Centre Informatique National de l'Enseignement Supérieur (CINES) depuis 2009. Elle contribue au fonctionnement de la plateforme d'archivage électronique PAC qui préserve, depuis 2006, les documents et données de l'ESR. Experte en pérennisation de l'information numérique, elle est membre de la CN171 de l'AFNOR, de la Commission Archives Électroniques de l'AAF et du comité de pilotage du SEDA. Elle participe également aux travaux du groupe PIN de l'association Aristote. Ses centres d'intérêt se portent sur la qualité et la certification des systèmes d'archivage électronique, les procédures d'archivage ainsi que les métadonnées de pérennisation. Depuis 2010, elle assure plusieurs formations sur ces sujets notamment pour la direction générale des patrimoines du Ministère de la Culture.

Depuis 1980, le CINES est un *data center* responsable de l'archivage de données produites par les établissements de l'ESR. Ces données doivent être validées, sélectionnées et documentées.

Lorsqu'on n'applique aucun traitement à des données numériques, elles peuvent se perdre en 10 ou 20 ans, voire 2 ou 3 ans, par exemple à cause d'un stockage défaillant, ou d'un support qui s'altère.

Les archives sont l'ensemble des documents, quels que soient leur date, leur lieu de conservation, leur forme, leur support, produits ou reçus par toute personne physique ou morale, et par tout service ou organisme public ou privé dans l'exercice de leur activité (Code du Patrimoine).

Le cycle de vie du document connait trois périodes :

- les archives courantes, nécessaires pour le traitement des affaires courantes ;
- les archives intermédiaires, d'un usage courant révolu, mais qui peuvent répondre à des besoins administratifs ou juridiques ponctuels ; elles posent des questions de facilité de recherche et d'intégrité ;
- les archives définitives, qui ont vocation à être conservées sur une durée illimitée pour des raisons historiques ou patrimoniales ; elles posent des questions de recherche et d'accès, d'intégrité, de compréhension du contenu.

Dans H2020, avec des projets en collaboration, le droit applicable est celui du coordinateur. En France les données relèvent de la législation sur les archives publiques. Une archive est publique lorsqu'elle est produite ou reçue par des services publics ou privés à mission de service public. Dans ce cadre, ne pas confondre « public » et « librement accessible » : une donnée publique n'est pas forcément publique, en fonction des règles d'applicabilité de la législation nationale.

La masse de données produites par la science est en pleine explosion (*Big Data*) et doit être accompagnée de réflexions sur la normalisation des formats, la présence d'une documentation et la qualité des données. Sans documentation, les données ne seront, à terme, plus

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

intelligibles. Sans suivi des formats, les données ne seront plus lisibles. La sélection en fin de période courante et de période intermédiaire rend le volume de données moins important : toutes les données n'ont pas vocation à être archivées, et ce tri s'opère en concertation étroite « archivistes/producteurs ».

Il importe de distinguer les différents niveau entre la sauvegarde (simple copie de sécurité destinée à poursuivre l'activité en cas d'incident) et l'archivage, qui implique la production de copies en sites distants, sur plusieurs supports afin de diversifier les stratégies d'archivage, avec respect de l'intégrité et de l'authenticité des données, veille et contrôle des formats, migration physique des supports, migration logicielle, métadonnées permettant l'intelligibilité des contenus dans une perspective de moyen/long terme. Sans archivage, on perd de l'accessibilité et de la compréhension de données.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425