



DIALOGU'IST - ATELIER #7

FAIRE VIVRE LES DONNÉES DE LA RECHERCHE

13 DÉCEMBRE 2018



CNRS – Réseau Renatis,
groupe de travail DIALOGU'IST



**UN 7^E ATELIER DANS LA
CONTINUITÉ**

UN 7E ATELIER DANS LA CONTINUITÉ

Des modalités de participation et de suivi

- 10 sites de visio conférence répartis sur le territoire
- 4 sessions : 1 état de l'art et 4 retours d'expérience
- 1 animateur pour tout le programme et 1 par centre
- 1 outil collaboratif pour le suivi, les questions et les notes

De la thématique

- Importance et enjeu de la gestion des données de la recherche dans un contexte de la Science Ouverte

Des événements

- Journées du groupe de travail inter-réseau « Atelier données » du 27 novembre
- La réunion annuelle RDA (Research Data Alliance) France du 5 décembre aux JNSO*
* Journées nationales de la Science Ouverte

FAIRE VIVRE LES DONNÉES DE LA RECHERCHE RETOURS D'EXPÉRIENCE SUR L'ORGANISATION, LA PRÉSERVATION ET LE PARTAGE

09h10-09h20 : Accueil et présentation du programme par Fabien Borget, enseignant-chercheur à Aix-Marseille Université

09h20-10h00 : Cadre politique et culturel de la Science Ouverte

- Francis André, Direction de l'Information Scientifique et Technique du CNRS

10h00-12h40 : Retours d'expérience

- Mise en œuvre de **Dataverse à l'Inra** par Esther Dzalé Yeumo, responsable de la DIST Inra
- Le **centre de données de géothermie Profonde**, un **exemple FAIR** par Marc Schaming, ingénieur de recherche à l'IPGS de Strasbourg, et Alice Frémand, Research Data Manager à l'EOST de Strasbourg
- Les **plans de gestion de données, archivage intermédiaire et Open Data** par Jean-Yves Nief, et Yonny Cardenas, ingénieurs de recherche, équipe de stockage du centre de calcul de l'IN2P3 Lyon
- **Mise en œuvre du principe FAIR appliqué aux données d'enquêtes qualitatives en SHS**, le cas de la banque d'enquêtes **BeQuali** par Sarah Cadorel, Ingénieur d'études Chargée des archives au CDSP, et Guillaume Garcia, Chargé d'études au CDSP, responsable de la banque d'enquête beQuali

12h40-12h50 : Clôture de l'atelier par Fabien Borget

FAIR, UNE NOUVELLE APPROCHE COLLABORATIVE

Des produits de la Science Ouverte et de leur partage et de leur circulation et réutilisation avec des recommandations

- Influe sur le signalement, le stockage, le traitement, les process, la réutilisation et l'évaluation
- Concerne le triptyque Chercheurs / Professionnels de l'IST / Informaticiens
- Favorise de nouvelles initiatives
- Fait émerger de nouveaux métiers et/ou compétences

→ UN CADRE POLITIQUE et un engagement des ACTEURS

LE CADRE POLITIQUE



PLAN NATIONAL
POUR LA SCIENCE OUVERTE

Prise en considération politique de la Science Ouverte
(intervention de Bernard Larrouturou aux JNSO)

- Impératif démocratique : réponses aux enjeux sociétaux, transparence de l'action publique de la Science Ouverte – **3 axes du plan national de la Science Ouverte**
- Impératif scientifique : meilleure science et meilleure efficacité via le partage
- Impératif éditorial : reprise en main de la diffusion de la Science

Ouverture des publications et ouverture des données

Le CoSO (Comité pour la Science Ouverte)

Au Cnrs, politique à venir

LE CADRE POLITIQUE

Plateforme OSPP (Open Science Policy Platform)

- 8 chantiers dont le Cloud EOSC, les FAIR Data, les modèles économiques des publications, les compétences, les sciences citoyennes, l'intégrité scientifiques, les mesures et métriques

L'appel d'Amsterdam (2016)

OpenAire

Différentes autres initiatives dont les identifiants persistants

RDA (Research Data Alliance)

- Nœuds nationaux dont le nœud Français



LES RETOURS D'EXPÉRIENCE

Des outils,
Des exemples de gestion de
données dans différents
domaines disciplinaires

DATAVERSE À L'INRA (DEPUIS FÉVRIER 2018)

Une réflexion politique institutionnelle de gestion des données (DataINRA)

- Intégré au contrat d'objectifs de l'Inra et dans la feuille de route de la DIST
- 1 lieu d'hébergement pour une capitalisation
- Le dépôt systématique (en complément de la publication) pour un outil alternatif aux solutions « éditeurs »

Le choix de l'outil dataverse pensé par et pour les Chercheurs

- Flexibilité de gestion des métadonnées
- Des conteneurs de données → création de (sous)collections
- Une gestion d'accès restreints possible

Un projet collaboratif multicompetences et copiloté

- DIST, DSI, Juriste, Chercheurs

DATAVERSE À L'INRA (DEPUIS FÉVRIER 2018)

- Automatisation de l'attribution des DOI via une API,
- Ouverture vers l'extérieur (OpenAire),
- MAJ possible
- Collaboration multicom pétence

- Licence CCO non adaptée (→ mettre CC-BY et etalab par défaut)
- Pas de vocabulaire contrôlé
- Identification via l'annuaire LDAP
- Intéropérabilité avec autres SI complexe
- Quelle évolution ?

Demain ?

- formation,
- consolidation technique,
- charte administrateur,
- certification des entrepôts,
- ouverture



LE CENTRE DE DONNÉES DE GÉOTHERMIE PROFONDE FAIR

Un contexte

- LABEX de géothermie profonde (univ. Strasbourg, Cnrs, ES - Electricité de Strasbourg), sociétés privées et financement mixtes)
- Une infrastructure de données spatiales (IDS GeOrchestra) : norme inspire iso 19139, données contrôlées, archivées
- Existence du noeud de l'infrastructure European Plate Observing System (EOPS)

Des données à traiter

- Données anciennes sur différents supports : les rassembler, les trier, les numériser et les documenter

Un process de FAIRisation

Facile à trouver : [catalogue cdgp](#) et plateforme EPOS

Accesible : DOI, protocole standard et système d'authentification

Interopérable : fichiers et métadonnées les plus riches possibles

Réutilisable : description la plus précise possible des données (dont auteurs et licences)

LE CENTRE DE DONNÉES DE GÉOTHERMIE PROFONDE FAIR

Un projet valorisé (post-doc, chercheurs ...) et à valoriser (démonstration, sensibilisation auprès des chercheurs...)

Un projet répondant à la Science Ouverte (DMP Opidor)

Un projet mutualisé (parti de ce qui existait)

Un projet collaboratif (avec les chercheurs devenant des « dataScientists », des juristes...)

Un projet avec des perspectives (poursuite sur les données d'hier et d'aujourd'hui)

Quelques difficultés sur l'indexation : quel vocabulaire contrôlé utiliser ?
(OpenThéséo, Inist, ...)

LES PLANS DE GESTION DE DONNÉES, ARCHIVAGE INTERMÉDIAIRE ET OPEN DATA À L'IN2P3

La problématique

- Comment conserver et stocker à long terme les données massives et de plus en plus nombreuses au centre de calcul sur une technologie évoluant constamment et rapidement ?
- Combien de temps les garder ?

La réponse : une réflexion permanente multicom pétence (IST, Chercheurs, Informaticiens et juristes) pour une politique de gestion de données et d'Open Data

- S'appuyer sur le DMP
- Faire l'inventaire des données
- Clarifier la politique des données entre le producteur et le fournisseur de services
- Améliorer le cycle de vie des données
- Faire des économies sur les espaces de stockages
- Sensibiliser les chercheurs sur la différence entre stockage et archivage (accroître la connaissance sur le cycle de vie des données)
- Décrire et documenter les données

MISE EN ŒUVRE DU PRINCIPE FAIR APPLIQUÉ AUX DONNÉES D'ENQUÊTES QUALITATIVES EN SHS LA BANQUE BEQUALI

Le contexte

- Une banque d'enquêtes à conserver et à valoriser
- Une opportunité de financement EquipEx
- Un besoin de contextualiser les enquêtes pour favoriser leur réutilisation

Le projet

- Une équipe multicom pétence : archivistes, Chercheurs (participation active) et informaticiens, qui s'est accrue en RH (de 2 à 7 à ce jour)
- Des partenariats (informatique et documentation)
- Un travail s'appuyant sur des réseaux
 - Comité Scientifique et Technique,
 - TGIR Progédo,
 - TGIR HumaNum,
 - Réseaux professionnels (Isore, Mate-SHS, archives de France, réseaux de laboratoires dont Archipolis disposant d'un catalogue dataverse)

MISE EN ŒUVRE DU PRINCIPE FAIR APPLIQUÉ AUX DONNÉES D'ENQUÊTES QUALITATIVES EN SHS LA BANQUE BEQUALI

La solution : un traitement en 7 étapes

- Collecte des données
- Classement et description
- Numérisation
- Contextualisation via des métadonnées riches (DDI, DC, Tei, ...) avec un vocabulaire contrôlé
- Identification (1 DOI / enquête)
- Diffusion sur bequali.fr (accès réservé pour certaines enquêtes à données personnelles et/ou sensibles)
- Archivage au CINES



Demain

- Renforcer la sensibilisation au potentiel de réutilisation (IA) et former les enseignants,
- Améliorer la collecte avec un DMP
- Obtenir une certification

The background of the slide is an abstract pattern of swirling, concentric lines in various shades of blue and grey, resembling a topographic map or a fluid flow visualization. The lines are irregular and organic in shape, creating a sense of movement and depth.

CONCLUSIONS

LE CADRE DE LA SCIENCE OUVERTE APPROPRIÉ PAR LES SCIENTIFIQUES

Le cadre européen FAIR nécessaire du fait de la diversité des données de la Recherche, de leurs distributions et de leurs formats

La prise en compte de la gestion des données de la Recherche (DMP) bien ancrée dans tous les appels d'offre ;
le DMP est le document central de l'Open Science

Nécessité de connaître de cycle de vie des données concernées et de s'appuyer sur une infrastructure permettant le stockage, l'archivage et le process des données → besoins de (in)formations

- Des outils existent : Dataverse cité dans 2 Rex

Nécessité de bien appréhender les aspects juridiques pour une bonne (ré)utilisation

- Le triptyque Chercheur/Informaticien/Pro IST devient un quadriptyque (ajout du juriste)

Les projets prennent en compte les différents axes politiques de la Science Ouverte

- Généraliser l'accès ouvert aux publication
- Structurer et ouvrir les données de la recherche
- dans une dynamique durable, européenne et internationale

MAIS il est nécessaire de tenir compte de la problématique du long terme et des processus validant qualitativement les données

Les données de la recherche représentent un immense potentiel dans toutes les disciplines, encore beaucoup de choses à inventer

UN ATELIER TRÈS ENRICHISSANT

Près de 80

participants

Un document à lire en complément

Rapport qui décline un plan d'action européen pour une gestion FAIR des données de recherche :

<https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283>

**Vers un prochain atelier Dialogu'IST en 2019
combinant
données, TDM et Intelligence artificielle**