

# **Lettre d'information GT Atelier Données n° 1 - Juin 2020**

Atelier DIALOGU'IST 9 le 9 juillet 2020 matin

« Rendre FAIR les données, mais quelles données préserver ? »

Après deux ateliers sur la FAIRisation des données, le Groupe de Travail Dialogu'IST du réseau Renatis propose un nouvel atelier sur la préservation des données.

Que faut-il préserver et que peut-on éliminer ? Quel est le cadre juridique qui sous-tend ces problématiques ? Quelles conditions techniques jouent un rôle et sont à prendre en considération ?

Différents intervenants issus de divers horizons (laboratoires de recherche, bibliothèques, archives) répondront à ces questions en présentant leurs retours d'expérience, après une brève remise en contexte.

Rendez-vous le 9 juillet 2020 dès 8h30 (accueil) pour une demi-journée d'exposés.

Compte tenu du contexte sanitaire, l'atelier se fera essentiellement en visioconférence, et sur quelques sites d'accueil dont la liste sera prochainement fournie (à confirmer). Pré-réservation recommandée à [dialoguist@services.cnrs.fr](mailto:dialoguist@services.cnrs.fr) source : <http://renatis.cnrs.fr/spip.php?article524>

---

Des SHS au traitement de données

Mardi 07-07-2020 de 14h00-16h00 (en visioconférence, lien à venir)

Le nouvel atelier Humanités numériques du GDR MaDICS (Masses de Données, Informations et Connaissances en Sciences) vise à contribuer au rapprochement des communautés de la science des données de celles des sciences humaines et sociales (SHS) impliquées dans les Humanités numériques. Il a pour objectif d'incorporer des éléments méthodologiques utilisés en SHS (par exemple, processus de constitution de corpus et de mise en relation de données, méthodologies d'analyse de corpus...) dans les processus informatiques (par exemple, intégration des données, navigation et visualisation des données...), et vice versa.

Pour engager le dialogue entre ces communautés scientifiques, l'atelier MADICS-HN organisera une première conversation le mardi 7 juillet 2020 en visioconférence dans le cadre du symposium annuel du GDR MaDICS. Cette première table ronde visera l'échange entre les communautés de la science des données et celles des sciences humaines et sociales autour de la notion de « donnée ».

Problématique

Depuis que les SHS intègrent des approches informatiques à leurs travaux, elles constituent des données selon des protocoles aussi standardisés et ouverts que possible afin que les données puissent être utilisées par les machines. Toutefois, les SHS gardent des réflexes dans la constitution de celles-ci qui visent avant tout des utilisateurs « humains ». L'exploration et l'exploitation des données par les machines requièrent alors des adaptations (filtrage, nettoyage, formatage...) supplémentaires. La table ronde discutera des attendus des machines en lien avec les processus méthodologiques utilisés par les SHS pour constituer leurs corpus et mettre en relation les données. Quelles sont les méthodologies des analyses de corpus attendues ou visées par les SHS ? Dans quelles mesures les processus informatiques contredisent ou accompagnent ces analyses (par exemple, intégration des données, navigation et visualisation des données) ?

L'objectif sera d'ouvrir le dialogue entre les chercheurs (dont les doctorants et post-doctorants) des différentes communautés et les ingénieurs de celles-ci à partir d'exemples ou d'expériences. Cette première rencontre concernera principalement les éléments de vocabulaire utilisés par chaque communauté, l'une des premières difficultés rencontrée dans tout projet de recherche pluridisciplinaire. L'exemple choisi pour la discussion concerne le mot « donnée ». La notion renvoie-t-elle bien à la même réalité dans les communautés des SHS et des SI ? Qu'est-ce qu'une donnée ? Parle-t-on de « donnée » ou de « données » ? Qu'est-ce qu'une « donnée propre » ? Utilisable par un ordinateur ? Quels sont les pré-requis des données exploitables automatiquement ?

La table ronde est ouverte à tous, les participants sont invités à venir avec des exemples concrets de jeux de données, de méthodes de collectes de données et d'expériences.

---

Présentation de l'outil DMP OPIDoR Les équipes de l'Inist vous présentent l'outil DMP OPIDoR, outil d'aide à la création en ligne de plans de gestion de données, et répondent à vos questions !

Programme Un plan de gestion de données (DMP) est un document qui spécifie quelles données sont collectées ou générées, comment celles-ci sont gérées, partagées et préservées pendant et après un projet. Les équipes de l'Inist vous présentent l'outil DMP OPIDoR, outil d'aide à la création en ligne de plans de gestion de données, et répondent à vos questions ! Après quelques rappels sur le Plan de Gestion des Données, le webinaire aura pour objectif de présenter l'outil DMP OPIDoR et abordera plus particulièrement les points suivants :

**Le choix du modèle de DMP dans l'outil**  
**Les fonctionnalités d'aide à la saisie**  
**La saisie des produits de recherche**  
**Les fonctionnalités collaboratives**

!! les deux premières sessions sont complètes !! inscription pour la dernière session  
-> [https://us02web.zoom.us/webinar/register/WN\\_jJFE8xXwSc6h5PXEEKKPehA](https://us02web.zoom.us/webinar/register/WN_jJFE8xXwSc6h5PXEEKKPehA)

source : <https://www.inist.fr/services/accompagner/webinaire/webinaire-dmp-opidor/>

---

## Objectifs

Concevoir la version française du manuel Open Science Training Handbook, de Foster Open Science  
Découvrir des méthodes pédagogiques pour former à la science ouverte  
S'initier à une méthode de travail collaborative et intensive

Dans le cadre de ce sprint édition et traduction, 2 types d'activités vous sont proposées :

éditer la version française produite par l'association HackYourResearch à partir d'une solution  
traduire de l'anglais vers le français les passages non traduits

Chaque participant · e sera amené.e à travailler sur les aspects édition et traduction. Le texte à traduire et à éditer a été sélectionné pour sa qualité. Il est disponible dans d'autres langues. La version produite par les participant · e · s sera à son tour éditée par une traductrice professionnelle.

Edit 12/05/2020 : le sprint se déroule intégralement en ligne

la partie concernant les données

<https://book.fosteropenscience.eu/en/02OpenScienceBasics/02OpenResearchDataAndMaterials.html>

---

The Symposium will be an innovative mixed virtual and in person event. As travel restrictions are likely to still be in place, the symposium offers virtual participation options for those of the international data community who will not be able to participate in person.

The International FAIR Convergence Symposium is organised by CODATA and GO FAIR on 22-23 October 2020 in Paris. This event will provide a unique forum for advancing international and cross-domain convergence around FAIR. The two internationally recognized organisations will convene a global community with an interest in combining data across domains for a host of research issues - including major global challenges, such as those relating to the SDGs or the COVID-19 pandemic. Outcomes will directly inform the CODATA Decadal Programme 'Making data work for cross-domain grand challenges' and the GO FAIR community's convergence towards bottom-up development of the Internet of FAIR data and services globally.

The outcomes will address fundamental issues, such as an improved alignment with parallel and emerging data initiatives around the world and lessons learned during the current crisis (virus outbreak) in relation to international research data sharing and the role of FAIR.

The approach in many sessions will be detailed and technical, resulting in recommendations, consensus for action and draft implementations.

Participation is open to ALL researchers, data experts, funders, and science policy makers. Programme

The Symposium programme in virtual and in person format will be announced early in September.

It will include keynotes and/or plenary sessions introducing each of the four themes, parallel sessions of various formats (paper sessions, workshops, interactive practical sessions and demos) and a plenary poster session. A plenary panel discussion will close the second day of the Symposium. See also paragraph on session types above. Important dates:

30 April: call for session proposals released

---

### Objectif

Le DFDR comblera une importante lacune dans l'infrastructure canadienne de recherche en fournissant une plateforme unique à partir de laquelle les données de recherche peuvent être chargées, organisées, préservées, explorées, citées et partagées.

Un outil de recherche fédéré fournira un point focal pour découvrir et accéder aux données de recherche canadiennes, tandis que la gamme de services fournis par DFDR aidera les chercheurs à stocker et à gérer leurs données, à préserver leur recherche pour une utilisation ultérieure, et à se conformer aux exigences de gestion de données des institutions et des organismes de financement.

### À propos

Portage, Calcul Canada (CC) et l'Association des bibliothèques de recherche du Canada (ABRC) collaborent pour fournir une plateforme fédérée pour la recherche et la gestion des données de recherche (GDR) numériques et la découverte. Ils se réjouissent d'annoncer que le service du dépôt fédéré de données de recherche (DFDR) a maintenant terminé la phase bêta et entrera en production limitée.

Ils regroupent beaucoup de monde <https://demo.frdr-dfdr.ca/discover/html/repository-list.html?lang=fr>

22 May: registration open

20 July: deadline for session proposals

17 August: accepted session proposals notified

21 August: deadline for posters and lightning talks submissions

4 September: programme announced

2 October: close of registration

22-23 October: International FAIR Convergence Symposium Organization

Convenors: CODATA, GO FAIR

Supporter / Local organiser: Ministry of Higher Education, Research and  
Innovation of the French Republic