

Webinaire n°2
autour du « Guide de bonnes pratiques de gestion des données de la recherche » :
la qualité des données

Date : Lundi 5 Juillet de 14H à 17H

Connexion : Le lien sera envoyé jeudi 1 Juillet

Programme

14:00-14:05 : Présentation du GT « Atelier Données » et du webinaire n°2

14:05-14:40 : Quel lien entre qualité et données ? (Alain Rivet, Cermav et Henri Valeins, RMSB)

La qualité des données est pour tous un élément capital en vue de leur diffusion et de leur réutilisation. Mais quels sont les éléments qui permettent de dire que des données sont de qualité ? En effet cette notion concerne à la fois leur qualité intrinsèque que celle de leurs métadonnées associées. En vous présentant les différents principes de la démarche qualité, nous allons essayer de vous montrer des éléments nécessaires pour donner confiance dans vos données.

14:40-15:15 : Processus de contrôle et qualification des données dans un système d'observation océanographique (Christine Coatanoan, Ingénieur Gestion de données au Sismer, Ifremer Brest.)

Les données océanographiques sont issues d'un grand nombre de systèmes d'observation et représentent une grande variété de paramètres. Pour s'assurer du bon fonctionnement global de ces systèmes d'observation et pour produire des informations fiables et standardisées, les données font l'objet de contrôles et de traitements et sont conservées dans des formats répondant à des critères de normalisation afin de promouvoir l'interopérabilité des bases de données, nationales et internationales.

15:15-15:30 : Pause

15:30-16:00 : Cycle de vie de la donnée ouverte de qualité (Geoffrey Aldebert, Etalab)

La qualité des données est un élément essentiel du succès de l'open data car l'ouverture des jeux de données en elle-même n'entraîne pas directement leur réutilisation. Ce constat s'explique notamment par les difficultés que rencontrent les ré-utilisateurs lorsqu'ils souhaitent s'approprier les données ouvertes, et par conséquent pointe un problème de qualité.

Dans cette présentation, nous aborderons les différentes étapes qui compose le cycle de vie de la donnée ouverte de qualité. De la fédération d'une communauté à la production de données, nous détaillerons notamment les différentes étapes dans la conception d'un schéma de donnée permettant d'assurer la qualité. Nous aurons également l'occasion de revenir sur l'écosystème (guides, outils, processus et bonnes pratiques) proposé par Etalab pour l'accompagnement des administrations dans la production de leurs données.

16:00-16:30 : Enjeux de la qualité des référentiels et des métadonnées pour la communauté scientifique : vocabulaires contrôlés, alignements pour une meilleure fairisation et répliquabilité d'un modèle de données - au travers de l'exemple des outils de Frantiq. (Blandine Nouvel, Centre Camille Jullianet Véronique Humbert, Archéologie des Sociétés Méditerranéennes)

La qualité des données dans Frantiq passe par plusieurs outils développés et maintenus par le réseau, ses adhérents, ses partenaires et ses prestataires. Elle est présente à toutes les échelles : dans le catalogue collectif indexé normalisé en UNIMARC et l'indexation des notices à partir du

thésaurus PACTOLS, un vocabulaire contrôlé avec respect des relations sémantiques. Tous les outils sont normalisés et "fairisés" grâce à des identifiants pérennes et à des alignements avec d'autres grands référentiels (IdRef, Wikidata...).

16:30-17:00 : Retour d'expérience en écologie végétale sur les étapes d'homogénéisation des données (Eric Garnier, chercheur CNRS au CEFE)

La majorité des jeux de données en écologie végétale sont de petite taille et sémantiquement hétérogènes. Leur réutilisation pour des objectifs de synthèse demande par conséquent un important travail d'homogénéisation afin de pouvoir conduire des analyses pertinentes. Cette présentation montrera les étapes qui ont été nécessaires pour préparer les jeux de données qui ont conduit à l'identification de deux dimensions majeures du fonctionnement des plantes.

17:00-17:05 : Clôture