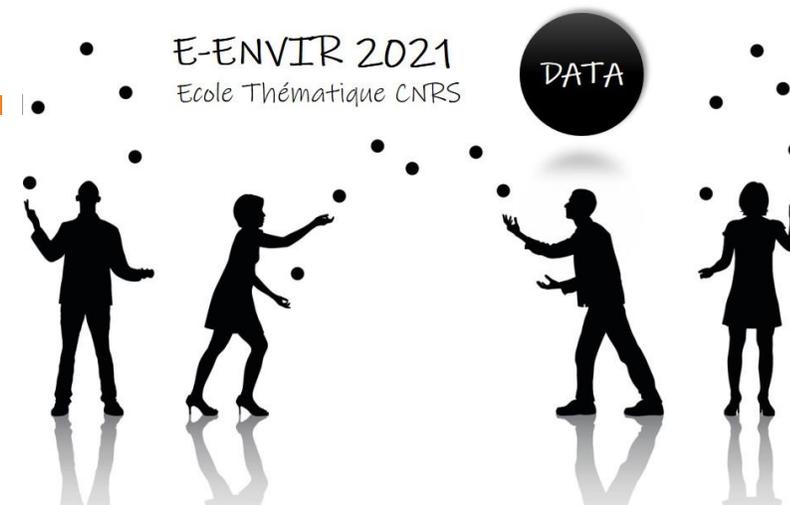


Stratégie & Recommandations à l'échelle du réseau eLTER France



MESSAGES A RETENIR AU DELÀ DE LA PRATIQUE

- La gestion des données est un processus complexe qui nécessite **des compétences et du temps**
- Les **coûts d'acquisition** des données **sont payés par de l'argent public** et il importe que ces données soient mises à la disposition de tous
- L'écosystème de gestion des données s'est complexifié
=> Essayez de trouver le bon niveau auquel vos données peuvent être gérées dans votre environnement plus ou moins proche
- Les IRs permettent **d'offrir de la visibilité** à vos données, mais - a priori - elles n'ont pas vocation à organiser et gérer vos données à votre place
- Un **Plan de Gestion de Données** permet de **clarifier le qui fait quoi**, de se poser les bonnes questions, de rechercher des solutions
- La **formation** et l'**accompagnement des doctorants** sur la question de la gestion des données sont des points clés !

LES ÉTAPES À RETENIR POUR LE PLAN DE GESTION DE DONNEES

A l'échelle d'une ZA, d'un observatoire, d'un site

- identifier les composantes de **son workflow sur la gestion des données** pour construire une image de son son environnement de gestion de données qui est souvent composé de gestionnaires de proximité ayant la capacité de vous aider et/ou d'accueillir vos données
- décrire le workflow (dont le vôtre) sur la gestion des données en **rédigeant un DMP** pour vous poser collectivement les bonnes questions, et identifier/aller chercher des ressources (humaines ou outils)
- le **DMP** aide à prendre du recul sur les pratiques en cours, et les points à améliorer, il **est évolutif**
- **assurer l'interopérabilité de ce Système d'Information** avec celui de votre IR
 - vocabulaire contrôlé, protocoles d'échanges, etc.

ROLE DES DIFFERENTS ACTEURS

- La construction d'un SI est un processus assez long (plusieurs années) qui s'appuie de préférence sur une équipe de permanents pour accompagner les chercheurs
- Remplir les BDD, les fiches de métadonnées est un travail ennuyeux peu valorisé, mais seuls les chercheurs peuvent le faire !
 - aller vers des outils/méthodes génériques, promus par les IRs
 - développer des outils ayant une valeur ajoutée pour les équipes scientifiques
- Le partage des outils et des méthodes facilite l'interopérabilité et minimise le travail

A RETENIR SUR L'INTEROPERABILITE

- **Interopérabilité technique**: à la charge des catalogues, entrepôts et systèmes d'information dans lesquels vous diffusez vos données.

Rend possible le moissonnage par des système tiers

Plus value:

- visibilité accrue de vos jeux de données, citation ()
- pas de duplication des efforts

- **Interopérabilité sémantique**: à la charge du producteur de données en décrivant ses données en utilisant des **vocabulaires contrôlés FAIR**

Plus value:

rend les données réutilisables par des tiers

A RETENIR POUR LA REUTILISATION

La documentation des données (provenance)

Documenter de manière détaillée les données que vous souhaitez partager: **se mettre à la place d'un utilisateur qui ne connaît rien à vos données**

(résumé succinct du protocole de collecte; d'expérimentation,..., sources de données en cas de réutilisation)

Les formats de données: recommandations GT format interpôle IR DATA TERRA

- Formats **ouverts, normalisés & auto-documentés, standards** du domaine.
- Contenant à la fois **données et métadonnées**
- Pour format ASCII, choix de l'**encoding de caractères = UTF-8** (et non ISO 8859-1 ou ISO 8859-15)

Formats recommandés:

- CSV
- NetCDF + convention pour les métadonnées

Les licences:

Mettre systématiquement une licence sur vos données précisant les conditions de réutilisation

Comment rendre vos données FAIR: à déterminer suivant votre position - l'appartenance à une IR ou pas



Geocatalogue
(labo, osu, instituts, ...)
Entrepôts au choix

Outils de type Geoflow ou
Metashark possible

Niveau 1 / à peine plus que d'habitude
(habitudes prises depuis e-envir21)

Domaine de IR



ZA Zones
Ateliers
LT SER FRANCE

Geocatalogue RZA
Entrepôt DataINDORES
Outils = geoflow, metashark



Outil = modèle pivot
In-situ OZCAR Theia

ICOS | Integrated
Carbon
Observation
System

SI ICOS France

Niveau 2 / plus fort : je FAIRise mes données avec l'appui des pole de données et selon
leurs recommandations et de manière compatible au SI de l'IR

Effort de structuration des données et des métadonnées

Quelle démarche recommandée: IR OZCAR

Les producteurs de données

- Assurent la collecte des données et leur validation scientifique
- Déterminent les règles d'accès et de réutilisation de leurs données.
- Fournissent des données qui respectent la loi RGPD et qui sont anonymisées si nécessaire.*

Les centres de données (OSUs ou institutions)

- Assurent la documentation des données (métadonnées)
- Assurent le stockage à moyen terme et parfois leur diffusion via leur propre système d'information
- Mettent en place le modèle pivot Theia/OZCAR ou les services permettant le transfert des informations nécessaires au SI Theia/OZCAR**

Le SI Theia/OZCAR

- Assure l'homogénéité de la présentation des données et leur visibilité sur un portail de données unique
- Assure le respect des règles fixées par les producteurs et la traçabilité des téléchargements de données
- A terme et selon l'avancement du développement de Data Terra RI, le SI Theia/OZCAR pourra proposer aux observatoires la possibilité d'obtenir des DOI de données et un archivage à long terme des données*

Quelle démarche recommandée: IR RZA

Déposez vos fiches de métadonnées dans vos **géocatalogues locaux** ou dans le géocatalogue du RZA
Mette un ou deux mot-clés rattachant votre fiche à une ZA ou au RZA "**LTSER-FR**"

Déposez vos données dans l'**entrepôt DataIndores de l'INEE** (géré par UMS BBEES)

Faire le choix entre Metashark et Geoflow ?

- Pour des données de biodiversité alimenter principalement le PNDB / GBIF et utiliser principalement la norme EML ⇒ Metashark
- Pour des données géolocalisées de tout domaines, alimenter principalement des géocatalogues (Geonetwork RZA) et utiliser la norme ISO19115 ⇒ Geoflow
- Pour publier dans des entrepôts (Zenodo/Dataverse + DOI) ⇒ Geoflow

Faut il vraiment faire un choix ?

pour toutes questions => lter-fr-bed@groupes.renater.fr

Non ! Cela va dépendre du contexte de travail, des projets, des pratiques, de la discipline, de l'accompagnement éventuel ...

⇒ D'abord définir vos besoins en terme de destination (qui sont parfois imposés/ fortement suggérés ...)

CAS PARTICULIER - IDENTIFIER DES SOLUTIONS

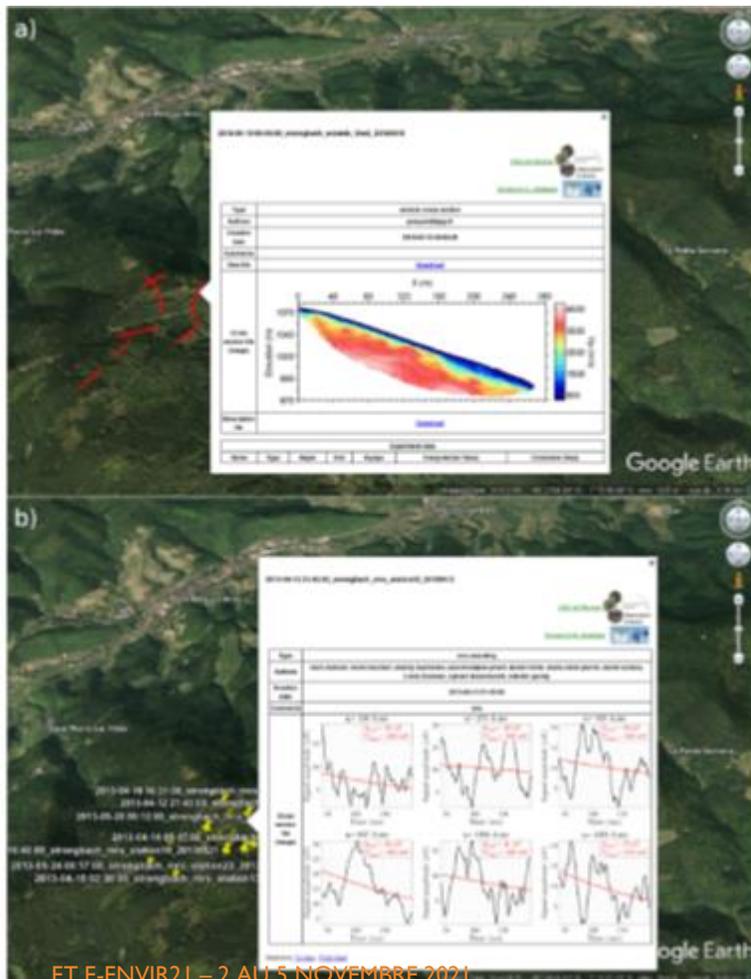
Si les solutions pour votre type de données n'existent pas à l'échelle locale, se rapprocher de vos homologues/pairs à l'échelle nationale, voire internationale

Exemples :

- Données hydrogéophysique ⇒ base de données spécialisée à Géosciences de Rennes
- Carottes sédimentaires ⇒ Cyber-Carothèque Nationale (échantillons) ou expérience ROZA (preuve de concept)
- Données en génétique ⇒ une communauté bien établie à l'international (GeneBank)
- Données en glaciologie ⇒ BDD internationale coordonnée par le WMO (World Meteorological Organisation)
- Mouvement des animaux ⇒ une communauté bien établie à l'internationale / MoveBank

Nouvelles opportunités et alternatives à la base SQL, les bases noSQL ⇒ exemple Cloud Environnemental

Base de données hydrogéophysiques



Développement initialement réalisés pour le SNO H+ pour bancariser différents types de données hydrogéophysiques: sismique, résistivité électrique, gravimétrie, etc..

Depuis 2018, l'IR OZCAR co-finance les développements et l'aide à la bancarisation des données hydrogéophysiques des observatoires de l'IR

Aperçu du module Google Earth pour le site du Strengbach avec a) les profils sismiques et b) les sondages par résonance magnétique. Au total, 10 profils sismiques et 16 sondages ont été insérés dans la base pour le site de Strengbach.

QUIZZ !!

<https://www.menti.com/9tg3ys1g66>

11 questions

RESSOURCES

1. Où trouver les informations ?

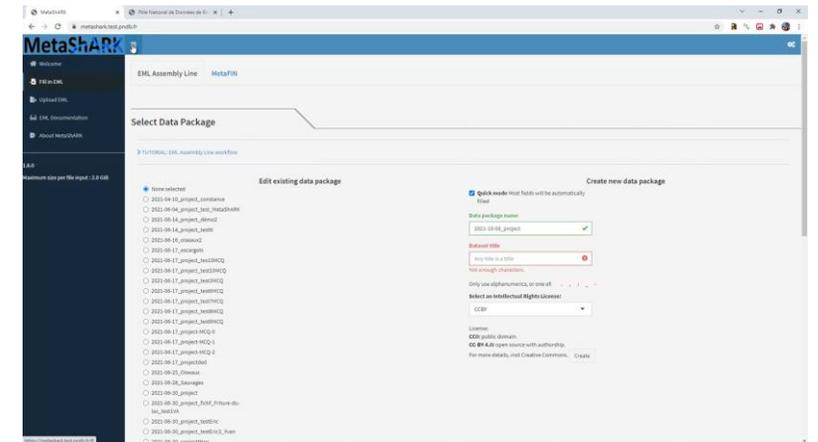
⇒ le site de l'e-envir21, documents téléchargeables ppt + vidéo-démo pour la pratique vue en TD sur la page "ressources"

1. Comment aller plus loin ?

⇒ consulter les sites DORANUM, une référence pour accéder à des contenus pédagogiques

1. Comment se tenir informer ?

⇒ RZA suivre et participer aux activités du GT4-BED
⇒ OZCAR newsletter THEIA/OZCAR
⇒ autres IRs ?



<https://data-access.cesgo.org/index.php/s/GSNKGSRInzXxyKq>



<https://doranum.fr/>

<https://doranum.fr/outils-services-gestion-partage-donnees-recherche/>

QUELLE SUITE DONNER À E-ENVIR 21 ?

Quels aspects de la formation pensez vous intéressant de présenter davantage ?

Quels autres aspects de la gestion des données vous semblent intéressant à proposer sous un format pratique / TD ?

BILAN E-ENVIR21 - VOTRE AVIS

Questionnaire d'évaluation de l'école e-envir21 (21 questions; anonymisé)

<https://enquetes-screen.msh-alpes.fr/index.php/843175?lang=fr>

A remplir par les **participants et les intervenants**
avant le **vendredi 12 novembre** (dans une semaine)

REMERCIEMENTS

cnrs

- CNRS soutien financier
- R. OMNES (conseillère formation, DR10) suivi de la formation
- P. RUPP (ICUBE), C. DIDIERJEAN (EDYTEM) gestion administrative et financière
- Le comité organisateur : BED/RZA, OZCAR/THEIA, UMS BBEES, PNDB, avec la participation de la MISHA
- Centre Saint-Thomas, hébergement

- Retrouvez les informations sur : <https://e-envir-21.sciencesconf.org/>



Maison Interuniversitaire des
Sciences de l'Homme
- Alsace -

