



# Retour d'expérience sur les écoles thématiques e-ENVIR2019 et 2021 sur la gestion des données en environnement

Isabelle Braud, INRAE, UR RiverLy, OZCAR, Lyon

Chloé Martin, CNRS, UAR BBEES, Paris

Isabelle Charpentier, CNRS, ICube, RZA, Strasbourg

Jean-François Le Galliard, CNRS, CEREEP ECOTRON, AnaEE, Paris

Séminaire Allenvi « Données des sciences du système Terre »  
Vendredi 24 juin 2022



# Plan de la présentation

1. L'école thématique e-ENVIR 2019: objectifs, contenu, public
2. Retour d'expérience
3. Pourquoi l'école thématique e-ENVIR 2021?
4. Retour d'expérience
5. Quels enseignements de ces deux écoles?
6. Perspectives et points d'attention



# e-ENVIR 2019: objectifs et contenus

## • Objectifs

- Etat des lieux des connaissances sur la gestion des données et la science ouverte chez les ingénieurs et scientifiques de l'INEE
- Former les acteurs aux principes généraux de l'open data en sciences de l'environnement (aspects conceptuels, juridiques et techniques)
- Partager une culture commune sur la production de données FAIR
- Informer sur les ressources et outils existants, en particulier les référentiels sémantiques



# e-ENVIR 2019: objectifs et contenus

## • Contenu

- Cadre juridique de la science ouverte
- Cycle de vie des données, principes FAIR et Plan de Gestion des Données
- Production de métadonnées standardisées
- Annotation sémantique pour les données en environnement, thésaurus existants et concepts d'ontologie
- Publication et mise à disposition des données (entrepôts)



# e-ENVIR 2019: objectifs et contenus

## • Public

- Inscriptions ouvertes aux ingénieurs et scientifiques de l'INEE et au-delà
- En majorité des ingénieurs ou scientifiques impliqués dans la gestion des données (34 participants, une dizaine d'intervenants)

## • Organisation

- Portée par AnaEE-France et l'UMS BBEES, avec les contributions des IR ACTRIS, OZCAR, ICOS, Data Terra
- Ecole thématique CNRS (formation permanente)
- 4 jours à Gif-sur-Yvette du 28 au 31 octobre 2019



# e-ENVIR 2019: retour d'expérience

## • Points positifs

- Une première vision globale des enjeux du partage et de l'interopérabilité des données en environnement avec des interventions de qualité
- Des retours d'expérience et un partage des expériences et pratiques des différentes IRs impliquées dans les domaines de l'environnement
- Formation des IT et chercheurs aux concepts d'annotation sémantique basée sur les ontologies
- La création d'une communauté de personnes intéressées et impliquées sur ces questions de gestion des données



# e-ENVIR 2019: retour d'expérience

## • Points à améliorer

- Un contenu qui est resté assez théorique avec trop peu de pratique et trop technique qui s'adressait plus particulièrement aux ingénieurs et techniciens
- Besoin d'outils et de pratiques plus ciblés sur les besoins des différentes communautés



# e-ENVIR 2019: retour d'expérience

## • Impacts de l'école

- A court terme:
  - L'identification du groupe d'intervenants pour e-ENVIR 2021
  - Identification des partenaires avec lesquels travailler pour publier leurs données
- A moyen/long terme:
  - Des formations spécifiques aux outils de la communauté AnaEE (2021 et 2022)
  - Volonté de refaire ce type d'école thématique plus ciblé sur certaines communautés
  - La création d'une communauté de personnes impliquées dans la gestion des données en environnement
  - Renforcement des structures existant déjà sur la gestion des données



# Pourquoi e-ENVIR 2021?

- e-ENVIR 2019 plutôt destinée à des spécialistes de la gestion des données et ressentie comme trop théorique et technique par certains participants
- Besoin d'outils pratiques pour aider les scientifiques à documenter et publier leurs données
- Une demande pour répondre aux critères de construction de l'IR européenne eLTER sur le partage des données
- Ecole portée par les IRs RZA et OZCAR et destinée prioritairement à leurs communautés, mais aussi ouverte à l'extérieur
- Programme tourné vers la prise en main d'outils correspondants aux pratiques des deux IRs
- Public visé plutôt scientifiques désirant publier et documenter leurs données



# e-ENVIR 2021: objectifs et contenu

## • Objectifs

- S'initier aux principes FAIR et aux niveaux de FAIRisation
- Être guidé par la pratique avec les recommandations actuelles des IRs OZCAR, RZA, PNDB et de l'UAR BBEES
- Traduire pour les non-spécialistes de la donnée les éléments techniques clés en piochant dans nos expériences et les nombreuses ressources existantes
- Se situer dans le paysage des offres en termes de gestion des données



# e-ENVIR 2021: objectifs et contenu

## • Contenu

- Préparation des jeux de données à la FAIRisation, plan de gestion des données
- Concepts clés des F & A du FAIR et des I & R du FAIR; spécificités de certaines données (notamment SHS) et les services des plateformes universitaires de données
- Présentation des outils utilisés dans les IRs OZCAR (modèle pivot Theia/OZCAR) et RZA (GeoFlow, MetaShark)
- TD pour utiliser pas à pas les outils proposés, avec des jeux de données mis à disposition et amenés par les participants et partager les bonnes pratiques
- Une vision stratégique sur le « comment je fais en pratique selon la nature de mes données »



# e-ENVIR 2021: objectifs et contenu

## • Public

- Public ciblé sur les IRs OZCAR et RZA
- En majorité des scientifiques impliqués dans les observatoires ou sites ateliers souhaitant partager leurs données (27 participants, une douzaine d'intervenants)

## • Organisation

- Portée par les IR RZA, OZCAR, PNDB et l'UAR BBEES
- École thématique CNRS/INEE (formation permanente)
- 4 jours à Strasbourg du 02 au 05 novembre 2021



# e-ENVIR 2021: retour d'expérience

## • Points positifs

- Pour les intervenants:
  - Une occasion de mieux comprendre les stratégies proposées dans les différentes IRs et les outils développés par chacune
  - Intervenants internes au comité d'organisation et bonne vision de ce que font les autres
- Pour les participants:
  - L'acquisition du B.A.-BA pour commencer à travailler avec les outils (cours et démo en TD)
  - La pratique : travailler avec son jeu de données

## • Points à améliorer

- La production de tutoriels de support pour accompagner les outils
- Un effort de traduction des termes/concepts employés encore à améliorer
- L'emboîtement fluide entre les différentes solutions, et entre les SI des IRs



# e-ENVIR 2021: retour d'expérience

## • Impacts de l'école

- A court terme: plusieurs jeux de données déposés sur les SI des IRs
- A moyen terme:
  - Peu de participants ont déjà réussi à mettre en pratique ce qui leur avait été montré (peu de jeux de données déposés, peu de plans de gestion des données)
  - Plusieurs témoignages d'une poursuite des dépôts si **accompagnement de proximité** par un.e spécialiste
  - Ouverture du portail InDoRES (INEE/BBEES) il y a quelques mois seulement, et outils maintenant disponibles pour déposer des jeux de données



# Quels enseignements de ces deux écoles?

- La gestion des données et l'open data: une question aux multiples facettes dans laquelle **ingénieurs spécialistes des données et chercheurs** doivent travailler de concert
  - L'implication des scientifiques dans la documentation et l'annotation des données est primordiale, mais difficile à obtenir: pas assez de temps disponible, activité trop peu valorisée. Les outils sont là mais ne suffisent pas!!!
- ⇒ Besoin d'accompagnement (ressources humaines) pour mettre les données en forme et les documenter pour éviter que les portails de données restent des coquilles vides
- La formation est importante mais doit éviter de rester trop théorique et doit apporter des solutions concrètes pour être appropriée par les scientifiques (outils, exemples, comment je fais en pratique)



## Quels enseignements de ces deux écoles?

- La multiplication de l'offre en termes d'infrastructure de données et d'entrepôts est source de perplexité et de confusion pour les scientifiques
  - ⇒ Besoin de clarifier le qui fait quoi et d'organiser les supports par niveau: quel support en local (OSUs, institutions, réseau des correspondants INEE, ateliers de la donnée), au niveau des IRs et pôles de données, au niveau national?
- Importance de travailler **en amont la documentation et l'annotation des données** pour une vraie réutilisation en particulier pluridisciplinaire (déposer sur un entrepôt de données n'est pas suffisant si c'est mal fait)
  - ⇒ Sensibiliser les scientifiques à l'importance de l'annotation des données, à l'usage de thésaurus, à l'intérêt de la sémantisation des données



# Perspectives et points d'attention

## • Perspectives

- Poursuivre les formations, actions de sensibilisation mais travail de longue haleine car il nécessite un changement de cultures et de pratiques
- Besoin de former et éduquer les plus jeunes (Master, doctorat) sur ces questions des données et de la science ouverte
- Encore des besoins d'harmoniser les formats selon les types de données (degrés de maturité différents selon les communautés)
- Généraliser les plans de gestion des données à tous les niveaux (observatoires, OSUs, IRs, etc..) et préciser les périmètres de chacun



# Perspectives et points d'attention

## • Points d'attention

- **Clarifier le paysage** pour les scientifiques (qui fait quoi) et qu'ils aient des points de contact locaux pour les orienter
- Veiller à ce que les **jeux de données déposés soient bien documentés** (vocabulaire, sémantique), sinon ils resteront inutilisables et l'interdisciplinarité ne pourra pas être mise en œuvre
- Importance de l'accompagnement au changement et **besoin de ressources humaines** pour cela (PEPR, autres?)
- Meilleure **reconnaissance** du travail sur les données dans les évaluations des personnels
- Besoin de **nouvelles compétences** (data curator, data manager, data steward) dans nos organismes



# Ressources et financements

- Soutien financier du CNRS/INEE (service formation permanente) et des IRs
- Ressources:
  - e-ENVIR 2019: <https://e-envir.sciencesconf.org/>
  - e-ENVIR 2021: <https://e-envir-21.sciencesconf.org/>



Merci pour votre attention

