



# RAPPORT ANNUEL SCIENTIFIQUE D'ACTIVITE 2018

## Infrastructures de recherche en environnement (IR) - Systèmes d'Observation et d'Expérimentation au long terme pour la Recherche en Environnement (SOERE)

### I Principales activités de l'IR / SOERE au cours de l'année 2018

En 2018 (première année de fonctionnement en IR), nous avons porté l'ensemble de nos actions sur deux actions transversales structurantes de l'IR : une animation scientifique transversale pour consolider notre projet scientifique, et une action transversale sur les bases de données. La première est portée par Sabrina Gaba et Camille Mazé (ZA Plaine & Val de Sèvre), la deuxième par Christine Plumejeaud (LIENSS, Univ. La Rochelle) et Cécile Pignol (ZA Alpes). Ces deux actions sont totalement disjointes, et sont donc présentées séparément dans le document ci-joint.

#### **Action transversale « Animation Scientifique »**

L'AT « Animation Scientifique » avait trois objectifs en 2018 :

**Faire émerger et partager la thématique de la résilience (adaptive resilience) et soutenabilité dans le RZA.** Trois actions sont proposées : l'organisation d'un symposium lors de la conférence Sfé<sup>2</sup> (Rennes, oct. 2018) sur « Resilience in Long-Term Social-Ecological Systems Research Sites », la participation active des membres du réseau des Zones Atelier (15) au workshop Garden Route (Afrique du Sud) regroupant des gestionnaires, des scientifiques et des décideurs publiques/politiques et la co-organisation d'une école thématique sur la résilience des socio-écosystèmes sur la Garden Route (Afrique du Sud).

**Mettre en place des projets transversaux à l'échelle du réseau.** Un projet inter-ZAs a été retenu à l'AàP du SOERE RECOTOX. Ce projet Néonet s'est déroulé sur la moitié des ZAs en 2018 et sera conduit en 2019 en incluant 3 autres ZAs. Nous avons également engagé une dynamique scientifique autour de la mise en place d'un réseau d'expérimentations socio-écologiques. En plus de réunions par visio-conférence, deux séminaires de travail ont été réalisés (Valpré, juin 2018 ; Rennes, oct 2018). Ces réunions ont permis d'identifier 3 problématiques de recherche qui conduiront à la mise en place d'expérimentations en 2019.

**Consolider les liens avec un ou plusieurs autres réseaux LTSER/résilience à l'international.** La présence de plusieurs membres du RZA et la co-organisation et participation d'étudiants à l'école d'été sur la Garden Route (Afrique du Sud) a permis de consolider des liens entre le RZA et le LTSER d'Afrique du Sud.

#### **Action transversale « Bancariser Ensemble les Données (BED) »**

Le projet BED a pour objectif d'améliorer la qualité et l'interopérabilité des données collectées et archivées par les chercheurs travaillant sur les Zones Ateliers en France. La démarche de BED est de travailler **DES LA COLLECTE** de l'information à sa documentation et traçabilité, afin de rendre les données *Findable Accessible Interoperable and Reusable* (FAIR). Nous insistons sur le fait qu'attendre que les données puissent être curées et rendues interopérables après assemblage dans d'hypothétiques bases de données est une illusion. Les membres des Zones Ateliers font presque tous partie de la « Long Tail of Science » et ne disposent pas de moyens informatiques pour traiter convenablement leurs données. Nous proposons donc des méthodes pour aider les chercheurs dans leurs tâches quotidiennes de collecte de données, afin de favoriser de bonnes pratiques avec des outils adéquats pour une documentation **UTILE** (au niveau des dictionnaires de variables) de leurs données, qui favorise la réutilisabilité de l'information mais aussi pour améliorer la qualité des données :

- traçabilité des échantillons grâce au projet COLLEC-SCIENCE,
- précision centimétrique à moindre coût des relevés terrain grâce au projet CENTIPEDE,
- respect des principes du RGPD avec ZAENQUETE,

- interopérabilité des modèles de données alignés sur Observation and Measurement et les ontologies de domaine en géologie par exemple grâce au projet ROZA,
- valorisation par la visualisation de données hétérogènes avec ELK.

## II Activité spécifique de l'IR/SOERE en 2018 incluant la prise en compte des recommandations du groupe Infrastructure de ALLENI

La vision que nous proposons s'est bâtie par le dialogue et une veille intensive dans le domaine de la gestion des données au cours des mois de l'année 2018. Nous avons ainsi été rencontrer les initiatives internationales autour de la gestion d'échantillon à la conférence *Research Data Alliance* à Berlin, nous nous inscrivons dans le nœud *RDA France* (rencontre à Paris le 05 dec. 2018) et nous avons été aux ateliers organisés par le CEFE à Montpellier et RBDD sur cette question en mars 2018. Christine s'implique dans le comité de pilotage RBDD, et est en contact avec l'IR OZCAR/THEIA sur ces questions. Les membres de BED sont aussi proches des OSU Theta, OASU, et OREME afin d'avoir une vision complète des problèmes et solutions envisageables.

Outre la veille scientifique et technologique, l'action de BED s'est focalisée sur 2 aspects :

- D'une part, **l'accompagnement et l'implication des chercheurs producteurs de données**, avec une réflexion sur l'usage de R pour la production facilitée de métadonnées, un atelier pour la formation à gestion d'échantillon et l'utilisation d'un ingénieur sous contrat pour un an (Oton Copi) qui a effectué des visites organisées sur les sites des Zones Ateliers sur des séjours d'au moins une semaine permettant d'aborder nombre de questions et problèmes.

Dans cette optique nous avons également démarré une action auprès des chercheurs en SHS afin de favoriser la bancarisation des données d'enquête dans le respect de la loi (LIL en France et RGPD au niveau européen). La journée du 28 Novembre a par exemple permis de souligner le rôle fondamental du contrat d'enquête entre enquêteur et enquêtés pour la collecte du consentement éclairé et de son archivage pérenne.

- D'autre part, la mise sur pied d'une **infrastructure de données minimale** avec le déploiement sur une machine virtuelle louée à l'INRA d'instances de COLLEC-SCIENCE, logiciel dédié à la gestion d'échantillons ainsi que d'un catalogue GeoNetwork moissonné par DEIMS, permettant de répondre à la directive INSPIRE, de mutualiser l'effort pour des ZAs parfois dépourvues de système de cataloguage, et de rendre nos données visibles au niveau européen et international. Nous avons étudié également la possibilité d'aider le développement de MDEdit2 (proposé par GeoGrandEst et la ZAEU) pour faciliter l'écriture de métadonnées de découverte. Et notre budget a aussi permis d'équiper deux Zones Ateliers d'un matériel d'impression et lecture d'étiquettes (ZA Loire et ZA EU), ainsi que de consommables. Nous avons souhaité également testé des étiquettes résistantes et collantes pour la congélation : marque CryoCool de Zebra. Des tests de mise en œuvre sur des terrains off-line ont été menés, sur la ZA PVS (V. Bretonnolle) et sur la ZAA (Jérôme Fort) nous permettant de tirer des conclusions pour l'amélioration du dispositif embarqué sur tablettes Windows ou raspberry Pis.
- Enfin une action de prospective active est menée à travers deux projets, ROZA et ELK pour **favoriser le croisement de données hétérogènes pour leur valorisation**. Dans la première, les données sont décrites et mappées sur les ontologies cadres (O&M) et de domaine (LIPD) tandis que dans la seconde, les données de capteurs sont croisées par le truchement de LogStash, et ElasticSearch et rendues visibles par Kibana. Ces deux actions permettent d'explorer des directions orthogonales pour apporter des solutions à l'ensemble des données déjà collectées dans le RZA, et permettent de distinguer des options différentes pour :
  - o les données dynamiques mais non véloces comme les données d'analyse de carottes sédimentaires (avec des questions essentielles de traçabilité dès le terrain) : ROZA

- les données véloces issues de capteurs en flux tendu mis en place par la ZATU : ELK

### III L'avancée ou les résultats des actions innovantes

#### Action transversale « Animation Scientifique »

- Article schéma conceptuel « Bretagnolle et al. 2018 »
- Symposium « Resilience in Long-Term Social-Ecological Systems Research Sites » qui a rassemblé 60 personnes dont la moitié ne faisait pas partie du RZA. Ceci a permis de faire connaître le réseau et de susciter l'intérêt de nouveaux scientifiques.
- Identification des problématiques de recherche qui seront appréhendées par une approche d'expérimentations socio-écologiques dans 4 à 10 ZAs en 2019.

### IV Eléments descriptifs et organisation de l'IR / SOERE (½ page max)

Infrastructure de Recherche Réseau des Zones Ateliers (RZA) ; [www.za-inee.org](http://www.za-inee.org)

**Directeur** : Vincent Bretagnolle [vincent.bretagnolle@cebc.cnrs.fr](mailto:vincent.bretagnolle@cebc.cnrs.fr)

CEBC CNRS, 79360 Beauvoir sur Niort - 05 49 09 78 17

#### Directeurs-adjoints

Hervé Fritz [herve.fritz@univ-lyon1.fr](mailto:herve.fritz@univ-lyon1.fr)

LBBE CNRS, UCB Lyon1, bâtiment Grégor Mendel, 43, boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex

Daniel Gilbert [daniel.gilbert@univ-fcomte.fr](mailto:daniel.gilbert@univ-fcomte.fr)

Laboratoire chrono-environnement, 4 place Tharradin, BP 71427, 25211 Montbéliard Cedex

Nicolas Lamouroux [nicolas.lamouroux@irstea.fr](mailto:nicolas.lamouroux@irstea.fr)

ZABR GRAY Campus Lyontech, La Doua, 66 boulevard Niels Bohr, bâtiment CEI, CS52132, 69603 Villeurbanne Cedex

#### Ingénieur administratif

Christine Schrive [christine.schrive@univ-tlse3.fr](mailto:christine.schrive@univ-tlse3.fr)

Ecofog, Campus agronomique, BP 316, 97310 Kourou – 05 94 32 93 00

#### – Structures de rattachement

CNRS-Institut Ecologie et Environnement

#### – Rappel du cadre d'activité, de la structure du système en réseau, des observatoires élémentaires

Les Zones Ateliers se focalisent autour d'une unité fonctionnelle (un fleuve et son bassin versant, les paysages agricoles ou urbains, la biodiversité, de l'Antarctique à l'Afrique Sub-saharienne, ou le littoral, ou encore la vie dans des environnements caractérisés par une irradiation chronique d'origine naturelle ou naturelle renforcée) et y développe une démarche spécifique en s'appuyant sur des observations et expérimentations sur des sites ateliers, pour y mener des recherches pluridisciplinaires sur le long terme. L'IR est le Réseau qui rassemble les 14 Zones Ateliers constituées auquel elle apporte

- une gouvernance identifiée, unifiée et effective,
- un pilotage stratégique et scientifique,
- un accès ouvert.

A terme, elle proposera également une politique de gestion et de diffusion des données.

### V Prospective d'évolution du projet en 2019(1/2 pages max)

Le bilan et les perspectives des deux actions transversales seront discutées en détail lors du prochain Conseil Scientifique de l'IR des Socio-écosystèmes, en décembre à Toulouse. A ce stade, nous pouvons cependant définir quelques grandes lignes directrices.

**Concernant l'Action Transversale « Animation scientifique » :**

- Conduite des expérimentations socio-écologiques autour de trois thématiques identifiées (Qualité des eaux, EcoHealth et Conservation de la biodiversité).
- Renforcer la formation des chercheurs et étudiants dans le réseau avec l'organisation d'une école thématique.
- Diffusion des recherches au travers d'articles scientifiques et de vidéos séminaire.

**Concernant la bancarisation des données**, nous souhaitons poursuivre l'animation et les actions entreprises, et nous nous sommes organisés en comité pour copiloter cette action. La vision que nous avons est de travailler **DES LA COLLECTE** de l'information à sa documentation et traçabilité, afin de rendre les données *Findable Accessible Interoperable and Reusable* (FAIR). Nous insistons sur le fait qu'attendre que les données puissent être curées et rendues interopérables après assemblage dans d'hypothétiques bases de données est une illusion. Les membres des Zones Ateliers font presque tous partie de la « *Long Tail of Science* » et ne disposent pas de moyens informatiques pour traiter convenablement leurs données. Nous proposons donc des méthodes pour aider les chercheurs dans leurs tâches quotidiennes de collecte de données, afin de favoriser de bonnes pratiques avec des outils adéquats pour une documentation **UTILE** (au niveau des dictionnaires de variables) de leurs données, qui favorise la réutilisabilité de l'information mais aussi pour améliorer la qualité des données :

- précision centimétrique à moindre coût des relevés terrain grâce au projet CENTIPEDE,
- traçabilité des échantillons grâce au projet COLLEC-SCIENCE,
- respect des principes du RGPD avec ZAENQUETE,
- interopérabilité des modèles de données alignés sur Observation and Measurement et les ontologies de domaine en géologie par exemple grâce au projet ROZA,
- valorisation par la visualisation de données hétérogènes avec ELK

Dans le détail, nos projets pourraient être les suivants :

- **COLLEC-SCIENCE - Gestion des échantillons et traçabilité des données associées**
  - **Réunion à Besançon sur 2 jours : mi Février 2019.** Dialogue avec tous les utilisateurs du dispositif en présence du développeur principal (Eric Quinton, IRSTEA) afin d'étudier toutes les pistes d'améliorations du logiciel (portabilité, prise en main, fonctionnalités) et d'organiser un comité de co-pilotage et une feuille de route pour les développements futurs.
  - **Actions de formation à COLLEC-SCIENCE** : au moins 2 ateliers à organiser sur 2 jours chacun. L'un au printemps 2019 sur Rhône-Alpes par exemple, l'autre à l'automne 2019 sur la Loire par exemple.
  - **Faciliter le déploiement de COLLEC embarqué sur une tablette Off-Line.** Il s'agit de développer une procédure basée sur Docker et résoudre des problèmes de compatibilité matérielle entre GPS et navigateur Web sur tablettes DELL Latitude 7212. Les prémices de la procédure sont publiés sur ce GITHUB :  
<https://github.com/jancelin/docker-collec/README.md>  
[https://github.com/jancelin/docker-collec/blob/master/docker\\_things.md](https://github.com/jancelin/docker-collec/blob/master/docker_things.md)
  - **Publier les échantillons sur le Web sémantique** : les échantillons peuvent se voir attribuer un DOI et être publiés avec leur description sur le Web Sémantique. Développer le programme dans le cadre d'un stage.
- **CENTIPEDE : améliorer la précision des relevés GPS au niveau centimétriques avec RTK**
  - **Tests d'antennes RTK** (achetées en 2018) sur la ZAPVS. Il s'agit de financer l'achat de smartphones résistants (type CrossCall action x 3) et de forfait 4G sur la zone, et de mener une collaboration avec les agriculteurs afin de déployer un réseau opérationnel sur 450 km<sup>2</sup>.
- **ZAENQUETE - Bonnes pratiques d'enquête (respect du RGPD/LIL):**

- Organisation d'une visio en janvier avec personnes présentes à l'atelier du 28 nov 2018 puis d'un **atelier** au printemps (1 journées à Paris) : méthodes d'anonymisation et d'archivage sécurisé des données.
- Action auprès de l'IRSTEA pour **ouvrir SYGADE en open-source** (négociation de moyens)
- Intégration d'un groupe SHS guide pratique CNIL ?
- **ROZA – méthodeS de publication sur le Web sémantique de données d'analyse**
- Organisation d'un atelier « **Choisir son vocabulaire** » du 11 au 13 mars 2019 à Chambéry, avec Barbara Magagna de DEIMS, Miled Rousset spécialiste d'OpenTheso, des membres du GDR SemanDiv et Sophie Aubin (INRA).
- Encourager la participation des chercheurs des Zones Atelier à l'école thématique **e-ENVIR** et à toutes les actions nationales expliquant la prise en main des outils du Web sémantique pour décrire leurs données.
- Participation à l'atelier de l'IR OZCAR/THEIA le 1/02/2019 sur la granularité des datasets et O&M
- exploiter les possibilités qu'offre **R et les packages *geometa*, *ows4R*** pour les chercheurs :
  - co-financer avec l'INRA l'avance (prestation) d'Emmanuel Blondel (FAO) sur certains packages (*geometa* pour produire des dictionnaires de données ISO 19110).
  - diffuser la pratique de création de métadonnées ISO 19110 (niveau dictionnaire de variable) avec R

Enfin nous **accompagnerons le déploiement d'entrepôts de données** au niveau des établissements, par exemple *Dataverse* pour l'INRA, le CIRAD et l'IRD. Il s'agit là de veiller à ce que leur mise en œuvre contribue à rendre les données pérennes et réutilisables, et donc à contraindre les workflows de curation, à favoriser l'articulation de ces entrepôts avec les infrastructures déjà existantes (en particulier les Infrastructures de Données Géographiques), afin d'éviter des saisies multiples de métadonnées. En particulier, le groupe de copilotage s'impliquera dans le **\*\*groupe DATAVERSE FRANCE\*\*** piloté par JC Desconnets afin de pouvoir influencer un peu sur la feuille de route des développements de cet outil pilotée par l'Université d'Harvard.