

# Outils R pour créer et mettre à jour des métadonnées - Organisation de la gestion d'échantillons.

## Objectifs:

1 - présenter une méthode, entièrement implémentée en R, pour **automatiser et standardiser l'écriture des métadonnées de découverte** et la publication des données associées (standards Dublin Core, OGC, EML, conventions CF, etc. pour les formats de métadonnées et WMS et WFS pour les protocoles d'accès). L'atelier montrera des exemples de codes qui utilisent uniquement des packages R pour écrire les métadonnées à partir de différents types de jeux de données (stockées dans des bases de données SQL type PostgreSQL, des fichiers tabulaires CSV, ou dans des fichiers NetCDF accessibles sur des serveurs Thredds) et des packages R pour les publier (basés sur les APIs de Geonetwork et Geoserver). Une chaîne de traitements (développée pour le projet H2020 BlueBridge) sera utilisée pour fournir des codes qui automatisent la mise à jour des métadonnées et des jeux de données associés. L'articulation de cette chaîne de traitement avec le système de bancarisation des données du Réseau des Zones Ateliers sera présentée. Repartir avec un ou plusieurs scripts adaptés à ses sources et jeux de données.

2 - partager un ensemble de pratiques autour de la **gestion des échantillons physiques** pour assurer leur traçabilité ainsi que celle des données qui leur sont associées. Illustration avec des cas d'étude en environnement. Réfléchir et repartir avec un schéma opérationnel et organisationnel adapté à son cas d'usage.

- **jour 1:** objectifs généraux, présentations des standards et des enjeux, et démonstrations simples d'utilisation des packages R,
- **jour 2:** manipulations (TPs) pour adapter les codes fournis en exemple aux sources de données des participants.
- **jour 3 (demi-journée):** considérations organisationnelles et pragmatiques quant à la gestion d'échantillons et la traçabilité des données associées dans les Zones Ateliers.

**Période et Lieu:** 17, 18, 19 septembre à la Rochelle / LIENS (salle de séminaire, 2 rue Olympe de Gouges, 17000 La Rochelle). <https://goo.gl/maps/Je3zLkTH8tQ2>

**Frais d'inscription:** gratuit, café offert, repas à la charge des participants.

**Matériel:** prendre son PC, le travail se fera en ligne sur l'infrastructure fournie par le projet BED (RStudio server).

**Participants:** Formulaire de pré-inscription : <https://survey.univ-lr.fr/index.php/142391?lang=fr>

## Liens vers les packages R:

- Standards OGC: [Geometa](#) (R tools to write, read & validate geographic metadata: OGC/ISO 19115, 19119, 19110, 19139), [Geonapi](#): R Interface to GeoNetwork API, [Geosapi](#): R interface to GeoServer REST API
- EML: package [EML](#)
- NetCDF: package [ncdf4](#)

## Lundi 17 Septembre

### Atelier métadonnées: bancariser ensemble les données dans le RZA

|               |  |  |                             |
|---------------|--|--|-----------------------------|
| 09:00         | Accueil  |  |                             |
| 09:30 - 10:00 | <b>Introduction / objectifs: infrastructure, workflow, sources de données, VRE et matériel pour le groupe:</b> | <i>Infrastructure BlueBridge: VRE, inscription et accès à RStudio Server<br/>Sources de données: OPeNDAP (Thredds..), SQL (PostGIS..), tabulaires (CSV)</i>  | <i>E. Blondel / J.Barde</i> |
| 10:00 - 10:30 | <b>Présentation générale du workflow</b>   | <i>Comment le configurer pour l'amorcer et l'utiliser pour faire des mises à jour des métadonnées...</i>   |                             |
|               | <b>Pause Café</b>  |  |                             |
| 11:00 - 11:30 | <b>Métadonnées statiques VS dynamiques: interfaces pour la saisie ou le calcul</b>                             | <i>Distinguer les éléments de métadonnées automatisables et ceux qui ne le sont pas..<br/>Fournir les informations dans un fichier CSV ou par requêtes SQL / OPeNDAP</i>                             |                             |
| 11:30 - 12:30 | <b>Standards de métadonnées: choisir des standards, comment lire les normes, mapping et packages R</b>         | <i>Exemple de la norme ISO 19115/110/119: diagrammes UML, schémas XML et exemples de métadonnées XML conformes. Autres exemples (EML..)</i>  |                             |
| 12:30 - 14:00 | <b>Pause Repas</b>   |  |                             |
| 14:00 - 15:00 | <b>TP: Utiliser les codes R et RStudio Server pour écrire une métadonnée simple..</b>                          | <i>Présentation d'échantillons de codes et l'environnement de compilation (IDE RStudio server)</i>   |                             |
| 15:00 - 15:30 | <b>TP: Introduction aux codes utilisés dans le workflow pour la gestion de (méta)donnée</b>                    | <i>Présentation générale: sans rentrer dans le détail des codes R, on présentera comment produire les métadonnées essentielles (titre, résumé, mots-clés, contacts...) à partir d'un fichier CSV</i> |                             |
| 15:30 - 16:00 | <i>Pause Café</i>  |  |                             |
| 16:00 - 17:30 | <b>TP: Introduction aux codes utilisés dans le workflow pour la gestion de (méta)donnée - suite</b>            | <i>Suite : on présentera comment produire les métadonnées essentielles à partir d'une table relationnelle interrogée en SQL et on ajoutera quelques métadonnées plus fines à titre d'exemple.</i>    |                             |
| 18:00         | <b>Fin de la journée</b>   |  |                             |

## Mardi 18 Septembre

### Atelier métadonnées: bancariser ensemble les données dans le RZA

|               |  |   |   |
|---------------|--|---|---|
| 09:00 - 09:30 | <b>Présentation détaillée du workflow</b>  | Détail sur le fichier de configuration du workflow (Json) et sur les différentes étapes et codes associés   | <i>E. Blondel / J.Barde</i>               |
| 09:30 - 10:00 | <b>Cadre général de la production de métadonnées dans le RZA</b>                             | Insertion du workflow R dans le paysage des Zones Ateliers. Présentation de l'articulation des différentes initiatives autour des métadonnées dans le RZA   | <i>C. Plumejeaud</i>                      |
| 10:30 - 11:00 | <b>Pause Café</b>  |   |   |
| 11:00 - 12:00 | <b>Ecrire une métadonnée 19115 sur un fichier NetCDF (accès OPeNDAP sur serveur Thredds)</b> | Adapter le workflow pour décrire des fichiers <b>NetCDF</b> ou NCML : accès OPeNDAP, description des jeux de données, créer des fichiers virtuels avec NCML pour modifier ou enrichir les métadonnées | <i>Julien Barde</i>                       |
| 12:00 - 13:00 | <b>Interagir en R avec Geonetwork et Geoserver</b>   | Publier les métadonnées et créer des Web Services associés (WMS/WFS...). Lien avec le catalogue des Zones Ateliers du RZA.  | <i>E. Blondel / C. Plumejeaud</i>         |
| 13:00 - 14:00 | <b>Pause Repas</b>   |   |   |
| 14:00 - 15:30 | <b>TP: gérer son propre workflow pour écrire une métadonnée 19115/19110</b>                  | <i>Décrire son jeu de données au format tabulaire CSV</i>   | <i>E. Blondel / J.Barde</i>               |
| 15:30 - 17:00 | <b>TP: gérer son propre workflow pour écrire une métadonnée 19115/19110</b>                  | <i>Décrire sa base relationnelle spatiale (MySQL, Postgis, etc.) et le catalogue de requêtes associé : connexion à la base, description des requêtes</i>  | <i>E. Blondel / J.Barde</i>               |
| 17:00 - 17:30 | <b>Pause Café</b>  |   |   |
| 17:30 - 18:30 | <b>TP: étendre les codes</b>   | <i>Utiliser les codes pour écrire d'autres métadonnées (EML, NetCDF, ou MetaDataCite). Lien avec le catalogue du GBIF par exemple.</i>  | <i>E. Blondel / J.Barde C. Plumejeaud</i> |
| 18:30         | <b>Fin de la journée</b>   |   |   |

## Mercredi 19 Septembre

### Atelier métadonnées: bancariser ensemble les données dans le RZA

|               |   |   |                                |
|---------------|---|---|--------------------------------|
| 08:30 - 09:00 | <b>Présentation détaillée du workflow de gestion des échantillons</b> | Organisation d'une collecte d'échantillons et des données associées : assurer la pérennisation des données et des | <i>C. Plumejeaud / O. Copi</i> |
|---------------|---|---|--------------------------------|

|               |   |  |                                |
|---------------|---|--|--------------------------------|
|               |   | échantillons - démonstration avec le matériel.   |                                |
| 09:00 - 09:30 | <b>Rédiger un protocole de collecte d'échantillons</b>                          | <i>TP : Check liste des items à documenter - exercice sur chaque cas particulier des participants</i>  | <i>C. Plumejeaud</i>           |
| 09:30 - 10:00 | <b>Discussions et retours sur les check-listes</b>                              | Échanges d'idées, trucs et astuces.<br>Convergence vers des préconisations pratiques.  | <i>C. Plumejeaud</i>           |
| 10:00 - 10:30 | <b><i>Pause Café</i></b>  |  |                                |
| 10:30 - 11:30 | <b>Manipuler collec-science : un exemple d'outil de gestion d'échantillons.</b> | <i>TP : Utiliser collec-science en ligne pour préparer sa collecte d'échantillons - fabriquer un gabarit d'étiquette, préparer un formulaire de saisie, etc.</i> | <i>C. Plumejeaud / O. Copi</i> |
| 11:30 - 12:00 | <b>Réflexion collective sur les usages et contraintes du terrain</b>            | Débat autour des améliorations à apporter, des solutions envisagées pour assurer des missions de collecte difficiles et off-line                                 | <i>C. Plumejeaud</i>           |
| 12:00 - 14:00 | <b><i>Fin de l'atelier - départ des participants</i></b>                        |  |                                |