



HAL
open science

Données et interopérabilité au sein de l'infrastructure de recherche AnaEE-France

Christian Pichot, Cécile Callou, Andre Chanzy, Aurélien Maire, Antoine Schellenberger

► To cite this version:

Christian Pichot, Cécile Callou, Andre Chanzy, Aurélien Maire, Antoine Schellenberger. Données et interopérabilité au sein de l'infrastructure de recherche AnaEE-France. Sist15: Séries interopérables et systèmes de traitement, Sep 2015, Marseille, France. 2015. hal-01604790

HAL Id: hal-01604790

<https://hal.science/hal-01604790>

Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Données et interopérabilité au sein de l'infrastructure de recherche AnaEE-France

Christian Pichot¹, Cécile Callou², André Chanzy³, Aurélien Maire², Antoine Schellenberger⁴

1 URFM, INRA, Site Agroparc, Avignon

2 US BBEES, CNRS, Museum d'Histoire Naturelle, Paris

3 UMR EMMAH, INRA/UAPV, Avignon

4 US Infosol, INRA, Orleans

Correspondance : achanzy@avignon.inra.fr

AnaEE France est une infrastructure de recherche dédiée à l'expérimentation sur les écosystèmes. L'infrastructure est portée par le CNRS-INEE et l'INRA et a reçu un financement des investissements d'avenir dans le cadre de l'AAP Infrastructures Nationales en Biologie et Santé. Par ailleurs, AnaEE-France est la composante française d'une infrastructure européenne qui est actuellement dans la phase préparatoire et inscrite sur la feuille de route ESFRI.

ANAEE-France réunit dans un réseau intégré un ensemble unique de plateformes expérimentales, de plateformes analytiques et de moyens de modélisation tous dédiés à la compréhension du fonctionnement des écosystèmes continentaux et de leur évolution. L'infrastructure s'articule autour de trois nœuds de services ouvrant l'accès à des dispositifs expérimentaux : la Très Grande Infrastructure de Recherche des Ecotrons, trois plateformes expérimentales en milieu semi-naturel et une vingtaine de sites naturels en France métropolitaine et en Guyane où sont conduites des expériences sur le long terme sur des forêts, des prairies, des cultures et des lacs. ANAEE-France offre également une instrumentation partagée et développe des moyens analytiques en microbiologie environnementale dédiés à la caractérisation fine de l'environnement et des micro-organismes. L'infrastructure propose une plateforme internet permettant de gérer l'accès aux dispositifs, de partager des bases de données, et d'accueillir des outils de modélisation. Ces outils numériques offrent de réels services pour réaliser des simulations de scénarios et constituent un moyen efficace de transfert des résultats scientifiques vers le monde socio-économique.

La politique des données d'AnaEE-France est de promouvoir la réutilisation des données acquises par les moyens de l'infrastructure. Nous distinguons les données de bases produites par les plateformes des données générées par les usagers des plateformes dans le cas d'expérimentation à façon. Pour promouvoir la réutilisation, deux services principaux ont été identifiés : i) une interface web permettant à travers des requêtes d'identifier les jeux de données et ii) l'interopérabilité entre les bases de données et les plateformes de modélisation. La mise en place de tels services nous amène à considérer plusieurs chantiers :

- Le flux de données des capteurs vers les bases de données. Le choix a été fait d'avoir deux types de système d'information, le premier dédié aux expérimentations à long terme et le second aux expérimentations à façon. L'acquisition et le traitement des mesures se fait au niveau de bases locales et les données sont ensuite transférées vers des systèmes d'information qui les archivent selon des types génériques adaptées à la nature des données.
- La gestion de la propriété intellectuelle. Ce point sensible a fait l'objet d'une attention particulière avec des droits par catégorie d'utilisateurs définis dans des chartes et régis par les systèmes d'information. Des outils permettant la reconnaissance des producteurs de données seront proposés (DOI).

- L'annotation des données par des métadonnées. Toutes les données sont annotées en s'appuyant sur un dictionnaire AnaEE-France et une ontologie. Dans la mesure du possible nous nous sommes appuyés sur les standards internationaux existants en intégrant des dictionnaires existants et en développant les ontologies selon l'approche OBOE largement utilisé dans le domaine de l'écologie. Des outils d'annotation sont en cours de développement pour fluidifier cette opération lors du dépôt des données dans les systèmes d'information.
- L'interopérabilité des bases de données et les plateformes de modélisation via des web services. Ceux –ci sont développés au niveau des plateformes de modélisation et visent à récupérer les données, les transformer lorsque cela est nécessaire et assurer la gestion des droits définis au niveau des SI.

L'objectif de la présentation sera donc de présenter le champ d'action couvert par l'infrastructure, l'ensemble de la démarche et les actions de développement qui sont nécessaires pour assurer la mise en œuvre de la politique de données.