



HAL
open science

Gestion des données dans AnaEE-France : enjeux et développements

André Chanzy, Christian Pichot, D Maurice, G Monet, A Schellenberger, Chloé Martin, Cecile Callou, Hélène Raynal, Nicolas Moitrier, François Lafolie

► To cite this version:

André Chanzy, Christian Pichot, D Maurice, G Monet, A Schellenberger, et al.. Gestion des données dans AnaEE-France : enjeux et développements. Enjeux sur le sol : les dispositifs de longue durée pour répondre aux questions d'aujourd'hui et de demain, Nov 2018, Versailles, France. hal-03591602

HAL Id: hal-03591602

<https://hal.inrae.fr/hal-03591602>

Submitted on 28 Feb 2022

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.

GESTION DES DONNEES DANS ANAEE-FRANCE : ENJEUX ET DEVELOPPEMENTS

André CHANZY et Christian Pichot

INRA – Avignon

(D. Maurice, G. Monet, A. Schellenberger, Chloé Martin, Cecile Callou, Hélène Raynal, Nicolas Moitrier, François Lafolie + nombreux CDD ...)

Versailles

15/11/2018

ANAEE FRANCE, UNE INFRASTRUCTURE DE RECHERCHE



- Un ensemble de services ouverts pour l'expérimentation sur les Ecosystèmes
- Un accès centralisé et harmonisé à ces services
- Nœud Français de l'infrastructure européenne
- Développer une excellence technique
- Gestion des données et diffusion



Plateformes expérimentales



Appui aux expérimentations

- Laboratoire d'analyse
- Dispositifs mobiles de mesure
- IRM du petit animal

Systèmes d'information et modélisation

- Portail d'accès aux ressources AnaEE
- Bases de données
- Plateformes de modélisation

ENJEUX

- Les données occupent une place de plus en plus prégnantes dans les processus de recherche
 - Accroissement dans les capacités d'observation
 - Accessibilité facilitée par les technologies du WEB
- La contextualisation des données (métadonnées, variables associées) est un élément capital pour les exploiter (corpus d'origine diverse)
- Les Infrastructures de recherche ont pour mission d'acquérir et partager des données de qualité (avec une grande diversité des types d'observation dans le cadre d'AnaEE)
- Interopérabilité (données, modèles)



ANAEE FRANCE DANS LE PAYSAGE DES INFRASTRUCTURES



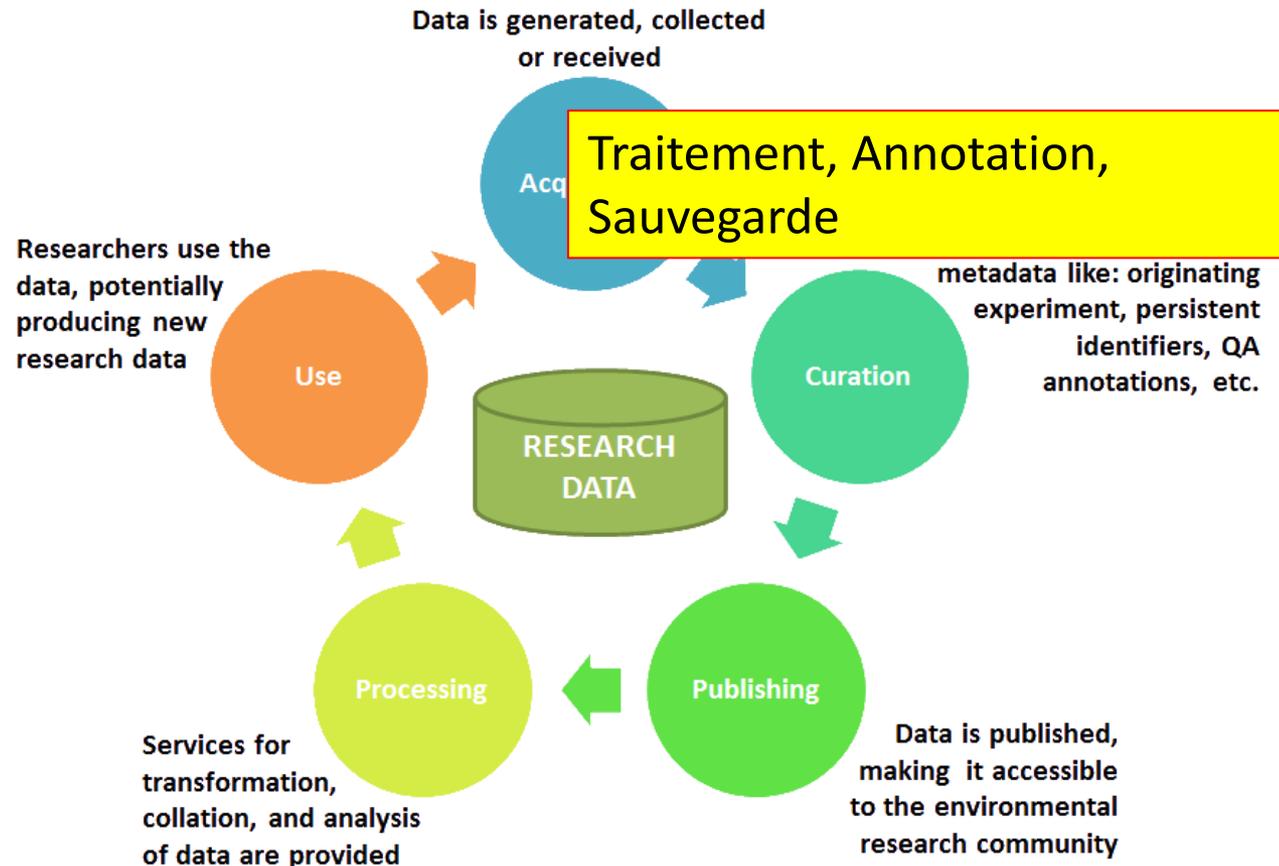
OZCAR



DONNÉES FAIR

- Findable
- Accessible
- Interoperable
- Reusable

Rôle des plateformes et usagers



Modélisation

Catalogage et Publication

→ Provide metadata

→ Give access to

→ Provide data

Providing information

INSPIRE EU directive
(facilities description, experimental
drivers, type of observations,
contact)

Access to data

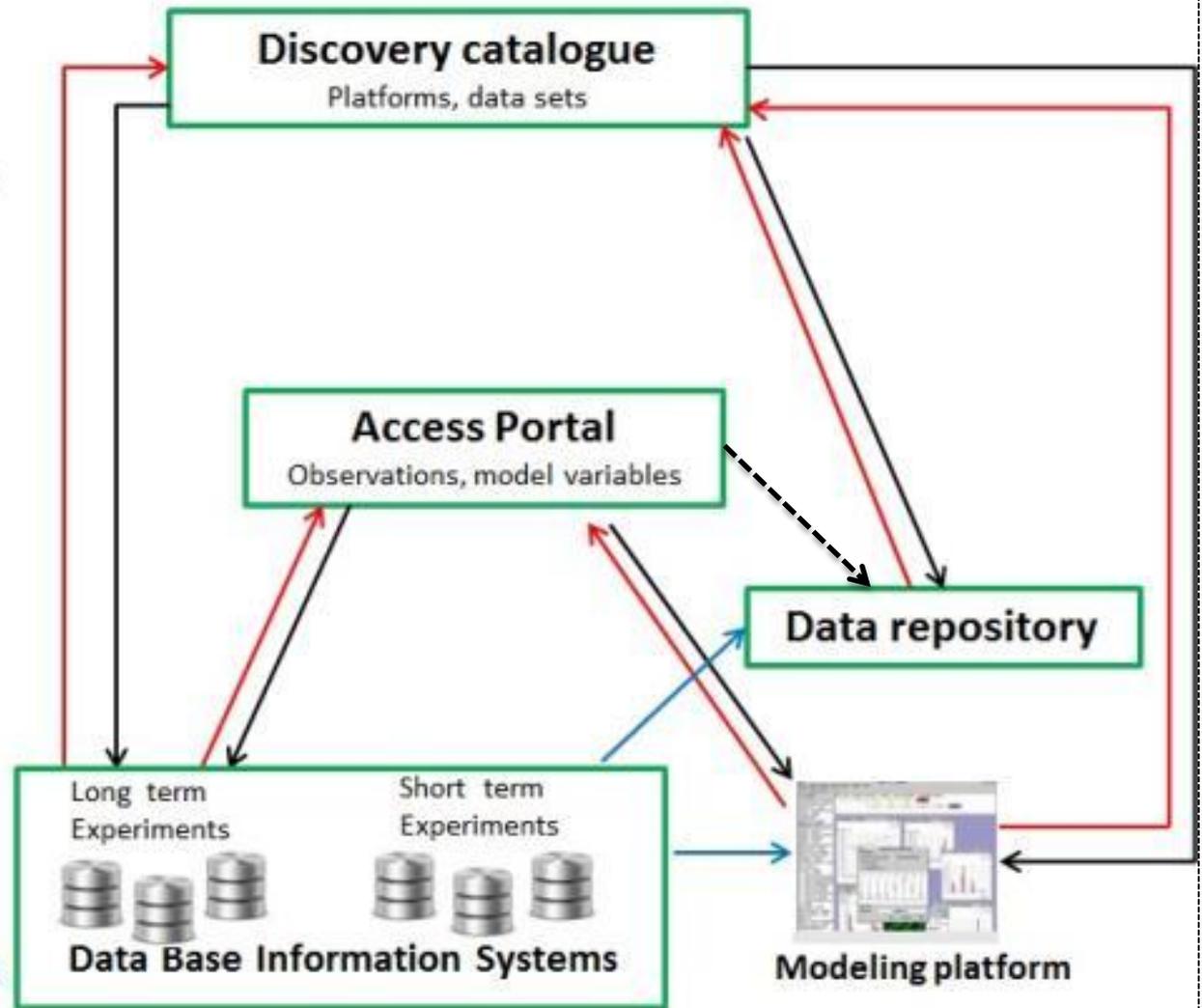
Semantic referential
(Data and model variable
description, provenance, license)

Data sets publication

DOI

Information System-DB

Shared development
Interoperability-webservices



LES SYSTEMES D'INFORMATION (CELLULE ECO-INFO)

https://si-acbb.inra.fr
Open since 2015
Home page

French / English



The screenshot shows the website interface with a blue header and a main content area. On the left, there are three vertical image galleries: a landscape with a road, a field with cows, and a weather station. The main text area contains a welcome message, a 'Comment accéder aux données ?' section with two buttons ('Se connecter' and 'Créer mes identifiants de connexion'), and a 'Pour en savoir plus sur l'observatoire :' section with a link to the SOERE website. A yellow callout box points to the language selection area in the top right, and another yellow callout box points to the 'Créer mes identifiants de connexion' button. The browser's address bar shows the URL https://si-acbb.inra.fr/acbb-web-1411.06/login.pl. The bottom of the page features a footer with 'Mentions légales', 'INRA', and 'Contacts'.



https://si-acbb.inra.fr
Open since 2015

SOERE ACBB (Agroecosystem) | X

Rechercher

Data retrieval | Data administration | Management of supplementary files | My account | Manage logs

Home

Welcome to the interface of data management information system of the SOERE ACBB.
The data are freely available subject to compliance with the [terms of use](#).
Some data are publicly accessible and can be obtained without special request. To access other data, you must perform one or more applications via the dedicated [links](#).
To see the data that are currently available in the database, you can use the dedicated link below in the page "Data Access".
You can find all the features of IS in this [manual](#).

Data retrieval

- Request access
- Extraction of gas flux and meteorological data
- Extraction of vegetation data.
- Data Synthesis
- Extraction of physical measurement of the soil data
- Extraction results
- Extraction of management practices data

Data administration

- Submit, publish, archive of experimental data
- Management of reference data
- Gestion des fichiers complémentaires.

Management of supplementary files.

- Management of user profiles
- Rights management of data sets
- Rights management of reference data

My account

- Edit Profile
- Manage rights applications

Manage logs

- Manage extraction logs

Data classification

Type of users

public
individual

Access

restricted
not restricted
public

Metadata

type of agroecosystems
experimental design
experimental treatments
plots characteristics
data types
Variables (name, description)
Units
etc.....

Cette interface vous donne un aperçu sur la répartition temporelle des données présentes en base et éventuellement vous permet de déceler des tendances sur les variables mesurées.

Arborescence des données disponibles

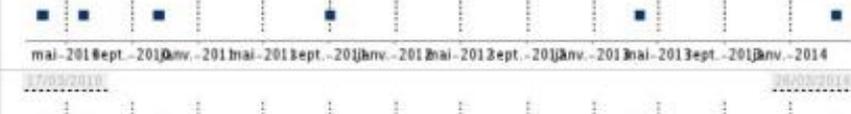
Sélectionnez un site et un type de données dans l'arborescence ci-dessous

- Estrées-Mons**
 - Semis du 17/03/2010 au 26/03/2014
 - Flux chambres du 06/09/2011 au 19/10/2014
 - Fertilisation du 29/03/2011 au 10/04/2013
 - Température du sol du 01/04/2010 au 31/12/2013
 - Tension de l'eau dans le sol du 01/04/2010 au 31/12/2013
 - Travail du sol du 17/03/2010 au 16/04/2014
 - Récolte et fauche du 19/07/2010 au 03/08/2011
 - Humidité volumique du sol du 01/04/2010 au 31/12/2013
 - Laqueuille**
 - Flux chambres du 22/02/2008 au 10/07/2013
 - Fertilisation du 30/04/2003 au 08/08/2013
 - Température du sol du 01/01/2003 au 31/12/2008
 - Pâturage du 10/05/2005 au 08/07/2011
 - Flux tours du 01/01/2002 au 31/12/2011
 - Météorologie semi-horaire du 01/01/2003 au 31/12/2009
 - Humidité volumique du sol du 01/01/2003 au 01/01/2009
 - Lusignan**
 - Semis du 20/04/2005 au 24/10/2012
 - Flux chambres du 22/11/2011 au 21/10/2012
 - Fertilisation du 11/07/2005 au 30/09/2013
 - Température du sol du 01/01/2006 au 01/01/2012
 - Tension de l'eau dans le sol du 01/01/2006 au 01/01/2012
 - Flux tours du 01/01/2009 au 01/01/2014
 - Météorologie semi-horaire du 01/01/2006 au 01/01/2013
 - Humidité volumique du sol du 01/01/2006 au 01/01/2012
 - Theix**
 - Fertilisation du 30/03/2006 au 09/08/2013
 - Température du sol du 31/05/2006 au 20/06/2012

Synchroniser

Votre sélection	
localisation	type de données
Estrées-Mons	Semis

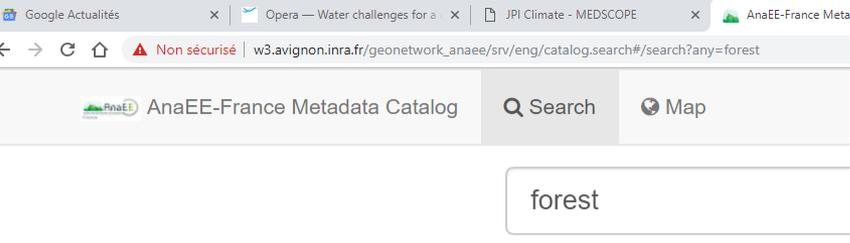
Liste des variables associées à votre sélection

Variable	Répartition temporelle des données disponibles
sem_espece	 <p>mai-2010 sept.-2010 janv.-2011 mai-2011 sept.-2011 janv.-2012 mai-2012 sept.-2012 janv.-2013 mai-2013 sept.-2013 janv.-2014</p>
sem_enrob	 <p>mai-2010 sept.-2010 janv.-2011 mai-2011 sept.-2011 janv.-2012 mai-2012 sept.-2012 janv.-2013 mai-2013 sept.-2013 janv.-2014</p>
sem_objectifs	 <p>mai-2010 sept.-2010 janv.-2011 mai-2011 sept.-2011 janv.-2012 mai-2012 sept.-2012 janv.-2013 mai-2013 sept.-2013 janv.-2014</p>
sem_depth	 <p>mai-2010 sept.-2010 janv.-2011 mai-2011 sept.-2011 janv.-2012 mai-2012 sept.-2012 janv.-2013 mai-2013 sept.-2013 janv.-2014</p>
sem_culture_annuelle	 <p>mai-2010 sept.-2010 janv.-2011 mai-2011 sept.-2011 janv.-2012 mai-2012 sept.-2012 janv.-2013 mai-2013 sept.-2013 janv.-2014</p>

CATALOGUE DÉCOUVERTE DES RESSOURCES



1) Conformité avec directive INSPIRE
2) Mots clés issus de thésaurus
3) Interopérabilité et moissonnage
(-> géocatalogue national)



géo catalogue Search Catalog Inquire Services Partenaires

Home Géocatalogue > Search > Results

WHAT WHERE

Sort by relevance 5 résultats / page

REFINE YOUR SEARCH

- Resource Type**
 - Datas (23)
- INSPIRE Themes**
 - Environmental monitoring facilities (21)
 - Soil (1)
 - Habitats and biotopes (1)
- Format**
 - Inapplicable (20)
 - IS query extraction files in csv (1)
 - C++ Application (1)
 - WAR (Web application ARchive) (1)
- Spatial Resolution**
 - 1/50000 (23)

Results 1 to 5 on 23

- AnaEE-France Infrastructure**
AnaEE-France is the french contribution to AnaEE which is a european project that is on the roadmap of ESFRI Committee (Biological and Medical Sciences) under preparation phase planned from 2011 to 2014
- UMS BBEES 3468 - AnaEE-F**
BBEES provides research units and researchers of the CNRS and the National Museum of Natural History with technical and scientific supports to structure, perpetuate or pool their databases. Its interventions...
- RMI for small animals**
RMI scan for Invertebrates and small Vertébrates (Fish, Amphibians, Reptiles, Birds)
- Information Systems for the "Observation and Experimentation Systems for Environmental Research" (SOERE)**
Observation and Experimentation Systems for Environmental Research (SOERE, A l'Envi label) are a network of sites devoted to the long term study of environment. Information Systems are developed by the INRA ...
- Ecotron IleDeFrance**
Together with the Ecotron de Montpellier, the Ecotron IleDeFrance is part of the distributed "Infrastructure de Recherche" (IR) managed and supported by CNRS and Ecole normale

Categories SOERE F-ORE-T : Breuil-Chenu Site

The former forest stand was a coppice with standards dominated by beech (*Fagus sylvatica* L.) and oak (*Quercus sessiliflora* Smith), and was clear-cut in 1976. After windrowing, the site was planted with monospecific plots. 2 bloks and 9 plots (0.10ha) Arnaud L

Categories SOERE F-ORE-T : B

The Bar... 33-ha Fontaine sessile Oak stand with a Hornbeam

CATALOGUE EXPLORATION DES DONNÉES

order by

order ascending

Ecosystems

Search

Clear

Ecosystems

Species

Start date

End date

Networks and experimental sites

Forest Ecosystem x

Type or select some options

Select an option

Select an option

Coffee-Flux x

Font-Blanche x

Coffee-Flux

Font-Blanche

play #

Résultats de la recherche

2 résultats

Export query results

Résultats de la recherche

Show 10 entries

Detail

+

+

Showing 1 to 2 of 2 entries

Infrastructure	Networks and experimental sites	Category of variables	Years interval	number of available data	Quicklook data from SI	Public
Soere Foret	Coffee-Flux	water cycle	[2009-2014]	385632	+ Soere Foret	
Soere Foret	Font-Blanche	water cycle	[2009-2016]	420768	+ Soere Foret	

First

Previous

1

Next

Last

Access

Soere Foret

Soere Foret

Soere Foret



ENJEUX DE L'ANNOTATION

- Pouvoir décrire la donnée pour qu'elle soit réutilisable :
 - Quoi, Comment, Où, Quand, Qui?
- Vocabulaire standard, mais gestion de la synonymie (requête non supervisée)
- Une description précise et rigoureuse interprétable par des machines
 - Recherche
 - Interopérabilité avec modèles et d'autres jeux de données

→ OUTILS SEMANTIQUES

- Vocabulaire contrôlé (thésaurus) => richesse sémantique et recherche sur termes voisins
- Ontologie → rigueur dans la description



THESAURUS ET ONTOLOGIE

Une diversité de noms dans la désignation d'un concept

Soil Moisture

Soil Water Content

Soil Water Amount

Soil Volumetric Water Content

Soil Gravimetric Water Content

Soil Humidity

THESAURUS va permettre de déterminer les synonymes et les concepts voisins

Contenu exact de la donnée

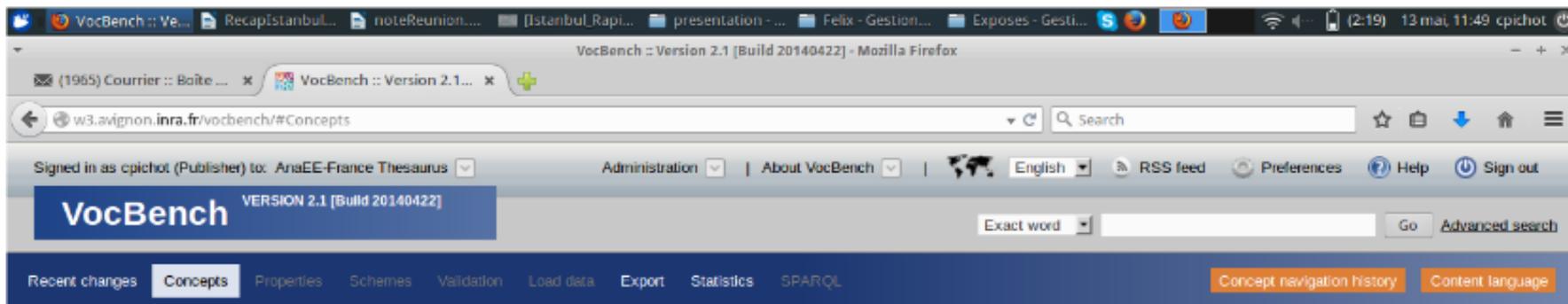
- Water Content, Water Amount => **quelle grandeur** exactement?
- **L'objet** auquel se rapporte l'information (surface, volume, couche)
- La représentativité **temporelle** (instantané (quand?), durée... (période, stat sur la période ..)
- **Où? Comment?**

ONTOLOGIE (description de la donnée)

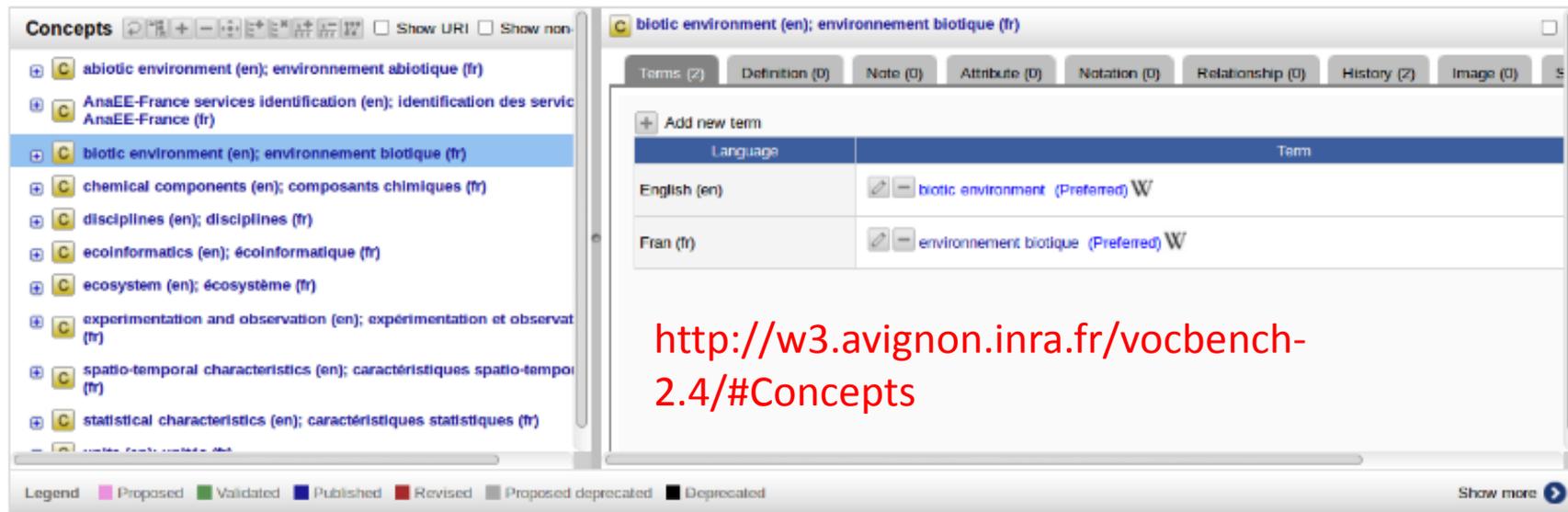


Tools for thesaurus management

VocBench (AML for alignments)



The screenshot shows the VocBench web application interface. The browser address bar displays "w3.avignon.inra.fr/vocbench/#Concepts". The page header includes the user "Signed in as cpichot (Publisher) to: AnaEE-France Thesaurus" and navigation links for "Administration", "About VocBench", "English", "RSS feed", "Preferences", "Help", and "Sign out". The main header features the "VocBench" logo and "VERSION 2.1 [Build 20140422]". A search bar is present with a dropdown menu set to "Exact word" and a "Go" button. Below the header, a navigation menu includes "Recent changes", "Concepts", "Properties", "Schemes", "Validation", "Load data", "Export", "Statistics", and "SPARQL". On the right side of the navigation menu, there are buttons for "Concept navigation history" and "Content language".



The screenshot displays the "Concepts" page in VocBench. On the left, a list of concepts is shown, including "abiotic environment (en); environnement abiotique (fr)", "AnaEE-France services identification (en); identification des services AnaEE-France (fr)", "biotic environment (en); environnement biotique (fr)", "chemical components (en); composants chimiques (fr)", "disciplines (en); disciplines (fr)", "ecoinformatics (en); écoinformatique (fr)", "ecosystem (en); écosystème (fr)", "experimentation and observation (en); expérimentation et observation (fr)", "spatio-temporal characteristics (en); caractéristiques spatio-temporelles (fr)", and "statistical characteristics (en); caractéristiques statistiques (fr)". A legend at the bottom indicates the status of concepts: Proposed (pink), Validated (green), Published (blue), Revised (red), Proposed deprecated (grey), and Deprecated (black). The main content area shows the details for the "biotic environment (en); environnement biotique (fr)" concept. It includes tabs for "Terms (2)", "Definition (0)", "Note (0)", "Attribute (0)", "Notation (0)", "Relationship (0)", "History (2)", and "Image (0)". The "Terms" tab is active, showing a table with two entries:

Language	Term
English (en)	biotic environment (Preferred) W
Fran (fr)	environnement biotique (Preferred) W

At the bottom of the page, there is a "Show more" button.

<http://w3.avignon.inra.fr/vocbench-2.4/#Concepts>

AnaEE Thesaurus

Summary Classes Properties Notes Mappings Widgets

Jump To:

- ⊖ abiotic environment
- ⊖ AnaEE France service identification and partners
- ⊖ biotic environment
 - ⊖ anatomy
 - ⊖ animal characteristics
 - ⊖ **biodiversity**
 - ⊖ biomass
 - ⊖ living organism (classification)
 - ⊖ microorganism characteristics
 - ⊖ phenology
 - ⊖ plant
- ⊖ chemical compound
- ⊖ disciplines
- ⊖ ecosystem
- ⊖ experimentation and observation
- ⊖ information system
- ⊖ model
- ⊖ spatio-temporal characteristics
- ⊖ statistical characteristics
- ⊖ units

Details	Visualization	Notes (0)	Class Mappings (2)	Access Class JSON
Preferred name				biodiversity
Definitions				1) Genetic diversity: the variation between individuals and between populations with species; species diversity: the different types of plants, animals and other life forms within a region; community or ecosystem diversity: the variety of habitats found within an area (grassland, marsh, and woodland for instance. 2) An umbrella term to describe collectively the variety and variability of nature. It encompasses three basic levels of organisation in living systems: the genetic, species, and ecosystem levels. Plant and animal species are the most commonly recognized units of biological diversity, thus public concern has been mainly devoted to conserving species diversity. (Source: WRES / GILP96)
ID				http://opendata.inra.fr/anaeeThes/c3_2393
broader				biotic environment
creator				demaodmin
definition				1) Genetic diversity: the variation between individuals and between populations with species; species diversity: the different types of plants, animals and other life forms within a region; community or ecosystem diversity: the variety of habitats found within an area (grassland, marsh, and woodland for instance. 2) An umbrella term to describe collectively the variety and variability of nature. It encompasses three basic levels of organisation in living systems: the genetic, species, and ecosystem levels. Plant and animal species are the most commonly recognized units of biological diversity, thus public concern has been mainly devoted to conserving species diversity. (Source: WRES / GILP96)
exactMatch				http://aims.fao.org/eos/agrovoc/c_33949 http://www.eionet.europa.eu/geme/concept/827
hasStatus				Published



2. Elaboration

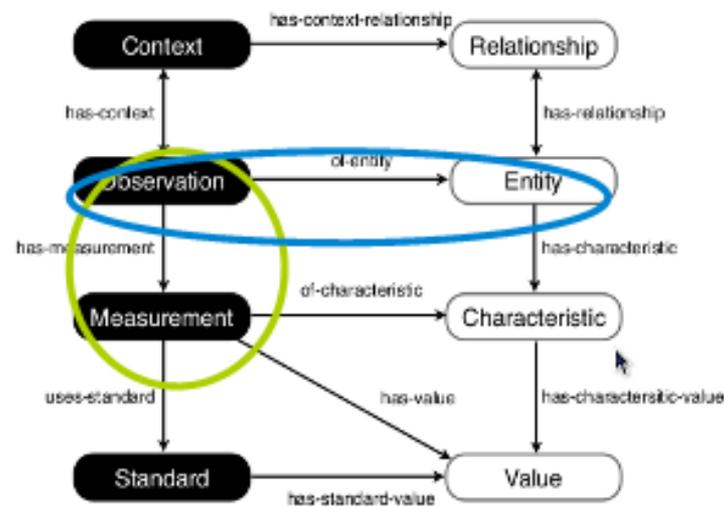
1. * basée sur OBOE, étendue à AnaEE-F
2. + autres ontologies de domaine

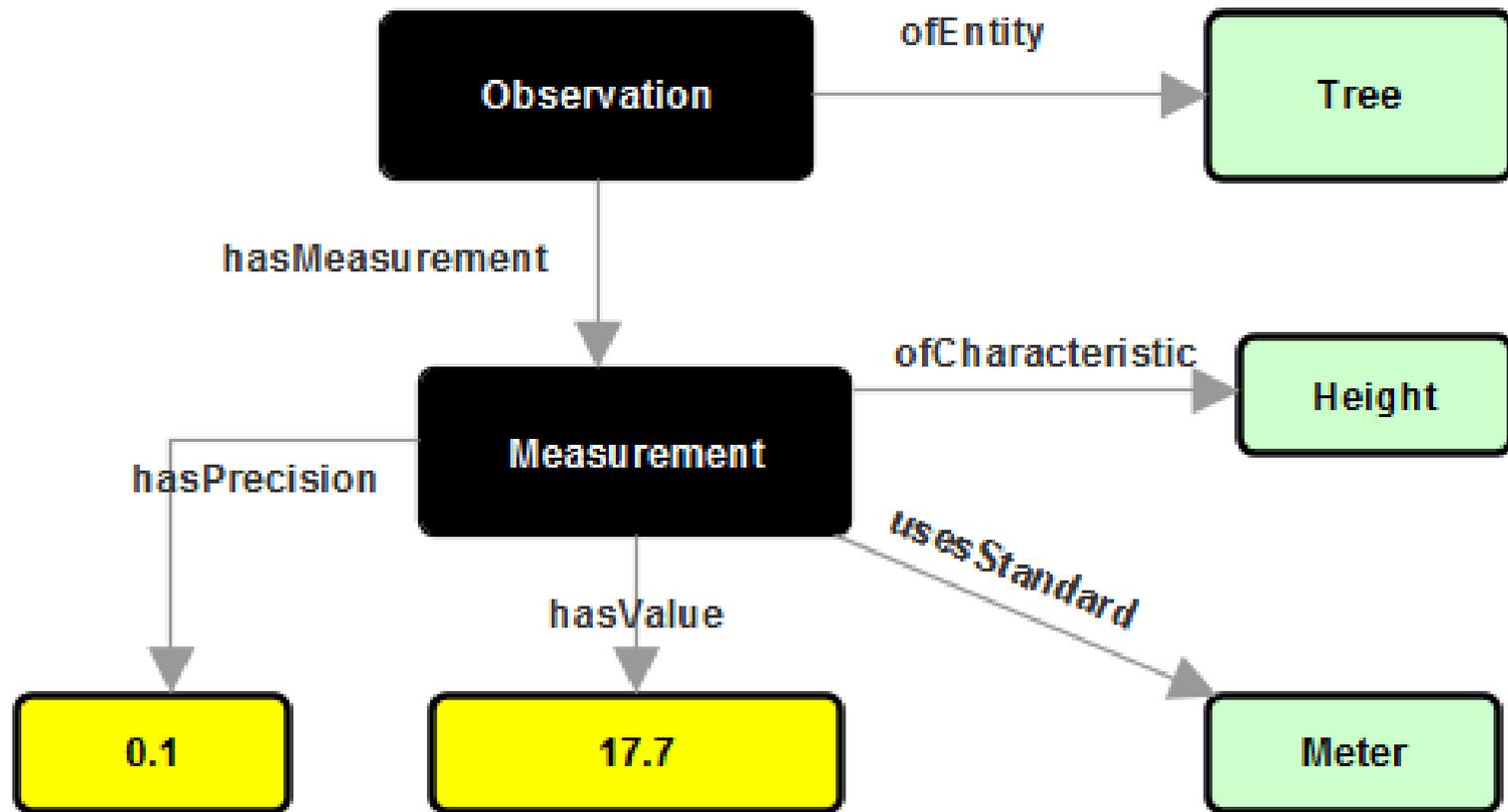
2. Méthodologie

1. * Géré dans l'outil "webprotégé"
2. * Analyse et modélisation sémantique des BDD et plateformes de modélisation

2. Utilisation

1. * Annotation par TOUS les Systèmes d'Information => triplestores
2. * Alimentation du portail d'accès (métadonnées de synthèse)
3. * Publication à venir (de l'ontologie) sous AgroPortal





Context

identification :

experimental data : measurement
localisation (plot, treatment, xy ...)

Simulated data :

Simulation identification

Spatial

characteristics :

Precise the spatial characteristics (not the location which is given in the context identification), when necessary of the compartment on which the measurement is done (eg soil layer requires to define the upper and bottom levels).

Quantity

identification :

Ontological Description of the quantity
(what the quantity represent without its spatial and temporal characteristic

Temporal

characteristics

Precise the temporal characteristics
(instant, interval, period ...)

Sensors/method

(SSN)



Graphe complet pour
l'azote ammoniacal
du SOERE OLA

=> Développement
d'un pipeline
d'annotation des BDD

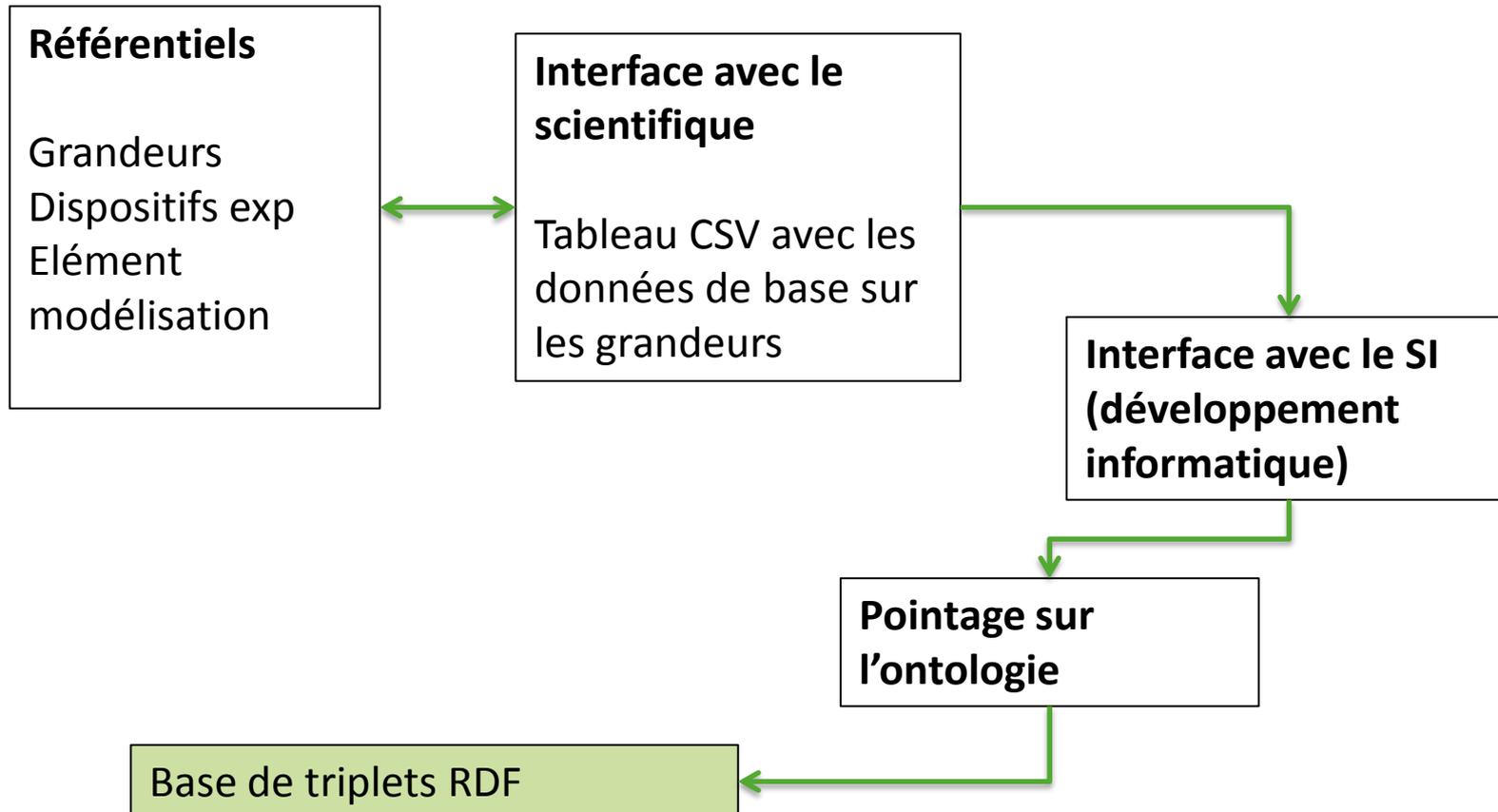


Nom Standard

AnaEE (thesaurus)	Cat. 1	Cat 2	Context	Context	Entité	Caractéristique	Prot.	Qualifi catifs	Unité
air relative humidity	meteorology	water cycle	atmosphère	air	water vapour	relative humidity			%
air relative humidity duration above 80%	meteorology	water cycle	atmosphère	air	water vapour	duration above 80%		daily cumulative	s
soil volumetric content	water soil physics	water cycle	soil layer; soil	soil material	water	volume per volume			m ³ .m ⁻³
potential evapotranspiration flux	meteorology	water cycle	crop	vegetation	potential evapotranspiration	volumetric flux	Penman	daily cumulative	m ³ .m ⁻²



Pipeline



DIFFUSION DES DONNEES

- Chartes des infrastructures institutionnelles, Contexte réglementaire (Loi Lemaire) → Open data
 - Ouverture large sans contrepartie financière,
 - Contraintes d'utilisation faible: Citation des sources.
 - Période d'embargo cadrée
 - Licences appropriées (~CC-BY)
- S'applique aux données générées par les plateformes et les usagers
- Publication des données avec un DOI
 - Quel périmètre?
 - Comment réactualiser un jeu de données dynamique (expérimentations long terme)?
 - Figurer les jeux de données, générer les fiches de métadonnées descriptives
 - Stocker le jeu de données dans un entrepôt de données



PUBLICATION

- Génération des jeux de données
 - Production de fichiers NetCDF (données plus métadonnées de contexte) en utilisant le pipe line sémantique pour générer les fichier.
 - Réflexion sur d'autres format de sortie
 - Génération des métadonnées (page d'accueil, documentation)
- Quel support de la diffusion des données (Systèmes d'information vs jeux de données publiés) ?
 - Via les SI (comment se référer à un DOI)
 - Via les entrepôts collectant les jeux de données publiés (on perd la richesse des interfaces d'exploration et d'extraction du SI). SI devient un outil de développement pour les partenaires impliqués dans la production des données
 - Via les deux (comment gère t-on les DOI)

CONCLUSIONS

- Ampleur du chantier (beaucoup plus que ce que l'on pouvait imaginer au départ)
 - Echelle Infrastructure vraiment pertinent (voir plus large)
 - Moyens humains(PIA)
- Communauté sur les données très active (Conceptualisation de la gestion des données, Standards ...)
 - Compréhension
 - Envisager un travail multidisciplinaire
 - Concevoir des interfaces entre producteurs de données et les spécialistes des données (pas seulement les informaticiens)
- L'intérêt des approches sémantiques
 - Supports très riches pour décrire les données, Rigueur
 - Souplesse pour évoluer et aligner des démarches séparées.

