

Des métadonnées pour améliorer la visibilité des BDD

Marie-Claude Quidoz (CEFE)

Atelier OSU : 03/07/2013

Métadonnée : définition

- Littéralement, ce sont des données sur les données
- Informations descriptives
- Répond aux questions : **qui quoi où quand et comment**
- Deux visions : celle du fournisseur et celle de l'utilisateur
 - Le fournisseur informe les potentiels utilisateurs des caractéristiques de ses données
 - L'utilisateur découvre des ressources et évalue leur pertinence pour ses besoins

Méthode 1 : définir ses propres champs

Qui sommes-nous ?
ENQUÊTE 2012-2013
Les bases
Actualités
Contacts
Offres d'emploi

Rechercher une base de données

Connexion
Login (identifiant de connexion au site):
Mot de passe:
[mot de passe oublié ?](#)
 Rester identifié quelques jours
Valider

Langues du site
français

Accueil du site > Les bases

Système d'Information Phénologique pour la Gestion et l'Etude des Changements Climatiques
Consulter la base ↗

Climatologie · Phénologie · PostgreSQL

Données phénologiques (flore et faune) du territoire français

Année de création : 2010
Copyrights : CNRS
Nombre d'entrées : plusieurs dizaines de milliers d'entrée
Accessibilité : Métadonnées accessibles en ligne
Publics concernés : Ecologues, climatologues (cinquantaine de personnes en France)
Personnels et son rôle : 1 responsable scientifique, 1 gestionnaire et 1 responsable technique
Présence d'un comité de pilotage ? : oui

PostgreSQL
SIG associé à la base ? : non
Statistiques associées à la base ? : non

Date de dernière modification de cette fiche : 16 février 2012

Phénologie



Responsable : Isabelle Chuine (UMR CNRS 5175 CEFE)

Contact : Isabelle Chuine - tel : 04 67 61 22 51
Isabelle.Chuine@cefe.cnrs.fr

Description sommaire : Suivi à l'échelle nationale de la phénologie de la faune et de la flore (dates des principaux événements qui rythment la vie a relation avec la saisonnalité du climat en différents sites chaque année)

Mots-clés thématiques : Phénologie - Faune - Flore

Région(s) concernée(s) : France

Date de début des mesures : 1370
Date de fin des mesures : En cours
Périodicité : Annuelle

Format des données : PostgreSQL
Mode de consultation : <http://www.gdr2968.cnrs.fr/>
Conditions d'accès : Limités aux membres du GDR jusqu'en 2010, ouverte à toute la communauté ensuite

Référence

Chuine I., P. You, N. Viovy, B. Seguin, V. Daux, et E. Le Roy Ladurie (2004) Grape ripening as an indicator of past climate. Nature 432:289-290.

TEXT SIZE - + Scroll To Top ▲

Méthode 2 : utiliser des standards

- Ne pas réinventer la roue (i.e. / champ à définir)
- Permettre l'exploitation par des agents logiciels
- Interopérabilité & Moissonnage
 - Facilité l'utilisation des données
 - Ne pas ressaisir plusieurs fois les mêmes informations
- **Mais quel standard prendre ?**

Extrait du Digital Curation Centre (DCC)

In this section

- Briefing Papers
- How-to Guides
- Developing RDM Services
- Curation Lifecycle Model
- Curation Reference Manual
- Policy and legal
- Data Management Plans
- Tools
- Case studies
- Repository audit and assessment
- Standards
 - Disciplinary Metadata
 - DIFFUSE
- Publications and presentations
- Roles
- Curation journals
- Informatics research
- External resources

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

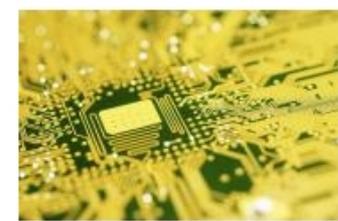
Search by Discipline



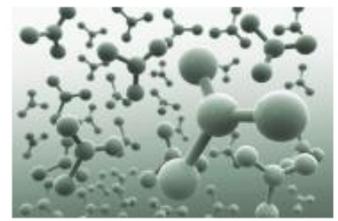
Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

Search by Resource Type

Earth Science

Extrait du Digital Curation Centre (DCC)

Geology Hydrogeology Palaeontology Molecular biology **Glaciology** Biogeography
Meteorology Minerology **Geoscience** Agricultural Economics Livestock
Ecology Marine Science Marine Biology Topography Genetics Hydrology
Entomology Genomics Botany Environmental Science Fish Farming Biochemistry
Planetary science Multi-disciplinary **Cartography** Oceanography
Hydrography Planning (Urban, Rural and Regional) **Geography** Soil Science
Climatology Agricultural Science

Metadata Standards

AgMES - Agricultural Metadata Element Set

A semantic standard for description, resource discovery, interoperability and data exchange for different types of agricultural information resources.

CF (Climate and Forecast) Metadata Conventions

A standard for climate and forecast "use metadata" that aims both to distinguish quantities (such as physical description, units, or prior processing) and to locate the data in space-time.

DIF - Directory Interchange Format

An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It includes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated.

FGDC/CSDGM - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata

BGDC policy states that non-Federally authored standards that are endorsed by the FGDC have the same status as FGDC developed standards. Since ISO 19115 and the associated standards are endorsed by the FGDC, federal agencies are encouraged to transition to ISO metadata as their agencies are able to do so.

ISO 19115

An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

Metafor CIM - Common Information Model

A model for describing numerical experiments carried out by the Earth system modelling community, the models they use, and the data they produce.



The Engineering Research Information Management (ERIM) project brought together the Innovative design and Manufacturing Research Centre (IdMRC) and UKOLN to explore effective data management, opportunities for and barriers to the re-use of engineering information, and requirements for the re-use of research data sets.

[Read more](#)

Biology

Extrait du Digital Curation Centre (DCC)

Climatology Ecology Plant physiology Multi-disciplinary Bioengineering
Palaeontology Proteomics Marine Biology Biochemistry Animal pathology
Metabolic biochemistry Bioinformatics Oceanography Soil Science Entomology Marine
Zoology Physiology Topography Genomics Genetics Plant pathology
Biomechanics Glaciology Marine Science Neuroscience Animal physiology
Zoology Botany Geology Hydrology Molecular biology Environmental
Science Meteorology Maritime Geography Cartography Biodiversity
Hydrogeology Geography Cell Biology Geoscience Hydrography
Biogeography Biomaterials

Metadata Standards

ABCD - Access to Biological Collection Data

A standard for the access to and exchange of primary biodiversity data, including specimens and observations.

Darwin Core

A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.

EML - Ecological Metadata Language

Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline.

ISA-Tab

A general purpose framework with which to capture and communicate metadata for data files from 'omics-based' experiments employing combinations of technologies.

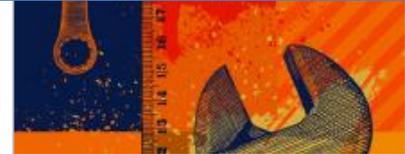
MIBBI - Minimum Information for Biological and Biomedical Investigations

A common portal to a group of checklists of Minimum Information in nearly 40 biological disciplines.

OME-XML - Open Microscopy Environment XML

A metadata standard and data file format for biological light microscopy data.

Repository-Developed Metadata Schemas



The catalogue was created to address the myriad challenges facing those who engage with data. Focusing on software and services that directly perform curation and management tasks, the catalogue splits the resources into five major categories, based on the intended user and the stage of the data lifecycle in which it will be most useful.

[Find out more](#)

Critères de choix

- Celui qui semble le mieux adapté pour décrire son jeu de données ?
- Celui qui est le plus utilisé dans sa communauté ?
- Celui qui possède l'environnement informatique le plus adapté à ses besoins (outils) ?
- Celui qui semble le moins contraignant pour les fournisseurs de données ?
- Celui qui possède le plus de module de transformation vers les autres standards ?
- ...

Des choix s'imposent ...

- Pour l'information géographique : ISO 19115
 - **LA référence**
 - Statut de norme internationale depuis 2003
 - Très proche des thèmes de la directive INSPIRE
 - Caractère modulaire et extensible
 - De nombreux systèmes de gestion de données géographiques basé sur ce standard
 - ...

Mais pour les données en écologie

- Pas de consensus
- Pourquoi ?
 - Les données sont plus diversifiées
 - Les utilisateurs sont plus « exigeants » (i.e. ont besoin d'en savoir plus pour exploiter les données)
 -
- Les plus usités : EML et Darwin Core

EML

- EML : Ecological Metadata Language
 - Norme de données élaborée par l'Écologie pour l'Écologie.
 - Initiative du National Center for Ecological Analysis and Synthesis (NCEAS)
 - Caractère modulaire et extensible
- Très utilisé aux USA ; peine à s'introduire en Europe
 - European Long-Term Ecosystem Research Network (LTER-Europe), Expeer, GBIF, ...
 - Par les zones Ateliers & OHM (INEE) - LTER

Darwin Core

- Préconisé par Taxonomic Databases Working Group (TDWG)
- Extension de Dublin Core pour le domaine de la biodiversité
- Orienté occurrences géographiques des organismes et l'existence physique des spécimens dans les collections biotiques.
- Utilisé par le GBIF

Des avantages et des inconvénients

- EML et Darwin Core sont complémentaires
 - EML : mieux adapté aux données complexes mais plus fastidieux à remplir et pas très courant en Europe
 - Darwin Core : adapté aux gestions de collections et assez simple à remplir
- Question récurrente :
Et pourquoi pas tout simplement utiliser l'ISO 19115 ?
(nos observations étant généralement géo référencées)
- **Dans la pratique, les parties « généralistes » des métadonnées étant compatibles entre elles et étant jugées suffisantes pour les fournisseurs pour faire connaître leur jeu de données, le choix se porte souvent sur l'ISO 19115 dans un premier temps.**

Et si le critère de choix était...

- Les standards des portails existants auprès des quels nous souhaitons faire connaître nos données (GBIF, INPN, LTER Europe,..voir l'OSU OREME)

Quelle est l'utilité de sélectionner un standard performant si aucun portail ne vient moissonner nos métadonnées ?

Projet CEFE

- Grandes diversités des jeux de données
 - 3 standard -> 3 prototypes
 - Darwin Core -> IPT -> GBIF
 - EML -> Metacat
 - ISO 19115 -> MDWEB (ou Geonetwork)
 - Avec 1 portail fédérateur ?
- Des interrogations cependant par rapport à la popularité d'EML en Europe