

Conserver et diffuser les données 3D

Les principes (FAIR) et les plateformes



Sarah Tournon-Valiente, plateforme Archeovision, Archéosciences Bordeaux

Olivier Marlet, CITERE LAT

Conserver et diffuser les données 3D

- Pourquoi FAIR
- Le contexte de la Science Ouverte
- Les bonnes pratiques de gestion des données au cours de Projets de recherche
- Les plateformes existantes

Pourquoi FAIR ?

Les principes FAIR



Findable / Trouvable

F1 Les données et les métadonnées sont identifiées par un identifiant global unique et pérenne.

F2 Les métadonnées décrivant les données sont riches.

F3 Les données et les métadonnées sont enregistrées et indexées dans un dispositif permettant de les rechercher.

F4 Les métadonnées spécifient l'identifiant de la donnée.

- *(Meta)data are assigned globally unique and persistent identifiers*
- *Data are described with rich metadata*
- *Metadata clearly and explicitly include the identifier of the data they describe*
- *(Meta)data are registered or indexed in a searchable resource*



DublinCore

Accessible

A1 Les données et les métadonnées sont accessibles par leur identifiant via un protocole de communication standardisé.

A1.1 Le protocole utilisé est ouvert, libre et peut être implémenté de manière universelle.

A1.2 Le protocole utilisé permet l'authentification et l'autorisation si besoin.

A2 Les métadonnées sont accessibles même quand les données ne le sont plus.

- *(Meta)data are retrievable by their identifier using a standardised communication protocol*

The protocol is open, free and universally implementable

The protocol allows for an authentication and authorisation when necessary

- *Metadata should be accessible even when the data is no longer available*



Interoperable

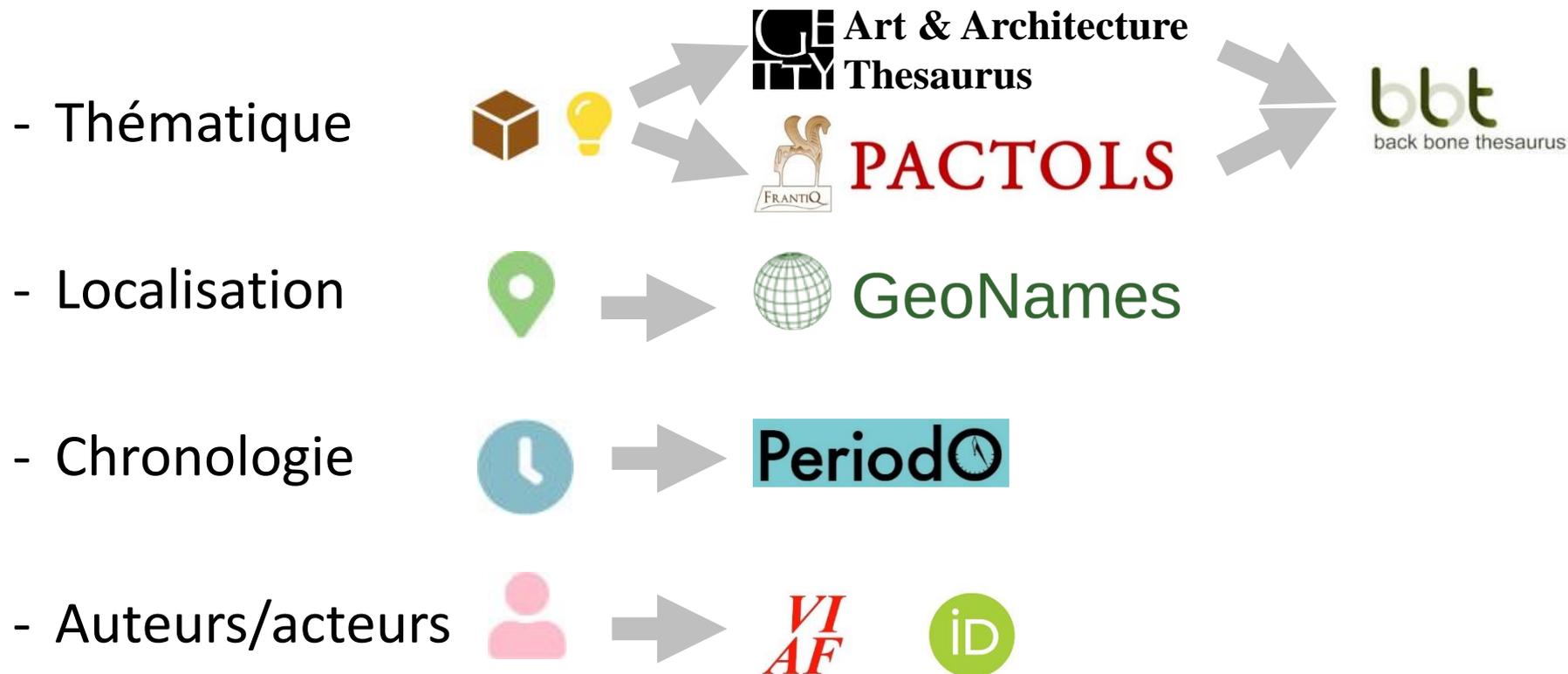
I1 Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.

I2 Les données et les métadonnées utilisent des vocabulaires qui respectent les principes FAIR.

I3 Les données et les métadonnées incluent des liens vers d'autres (méta)données.

- *(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation*
- *(Meta)data use vocabularies that follow the FAIR principles*
- *(Meta)data include qualified references to other (meta)data*

Interopérabilité : lier les concepts communs



Utiliser des thesaurus ou des référentiels standards

Référencer une URI plutôt qu'une chaîne de caractères (dépendante d'une langue)

Référentiels partagés par une communauté internationale (multilinguisme)

PeriodO + Geonames

PeriodO

PeriodO Current backend: Canonical [switch] Sign in Menu

Browse by: Period Collection

Periods

Viewing 26 - 50 of 660

Show periods at a time.

Previous ... Next

| ▲ Label | Earliest start | Latest stop |
|--|--|-------------|
| Bronze Age | | |
| Permalink: http://n2t.net/ark:/99152/p0qhb665rjw | | |
| In collection: ARIADNE Consortium. ARIADNE Data Collection. 2015. | | |
| Original label | Bronze Age (eng-latn) | |
| Spatial coverage des... | Poland | |
| Spatial coverage | Poland | |
| Start | 3800 BP (2000) (ISO value: -1800) | |
| Stop | 2401 BP (2000) (ISO value: -401) | |
| Notes in source | ARENA1/ADS | |
| Editorial notes | Derived from http://n2t.net/ark:/99152/p0zj6g8wft6 | |

GeoNames

Plan Satellite

Layers

blason Esvres ca. 66 m

ADM4 fourth-order administrative division **6433019**

France ^{FR} » Centre-Val de Loire ²⁵ » Indre and Loire ⁴⁷ » Tours ³⁷² » Esvres ³⁷¹⁰⁴

population : 5849
47.28444, 0.78611

N 47°17'04" E 0°47'10"

geotree .kml .rdf

Google

Imagerie ©2019 TerraMetrics | 2 km

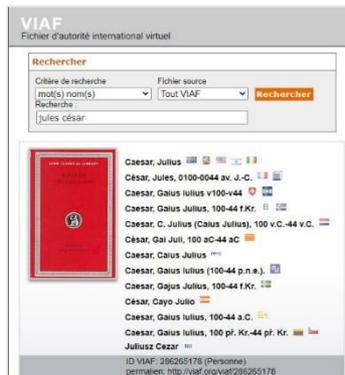
Conditions d'utilisation | Signaler une erreur cartographique

Identifiant unique (et pérenne) = non lié à une langue + fait référence à un concept

Auteurs : identifiant « unique » ?

Jules César

VIAF



VIAF
Fichier d'autorité international virtuel

Rechercher

Critère de recherche
mot(s) nom(s) Tout VIAF Rechercher

Recherche
jules césar

Caesar, Julius
César, Jules, 0100-0044 av. J.-C.
Caesar, Gaius Iulius v100-44
Caesar, Gaius Iulius, 100-44 f.Kr.
Caesar, C. Julius (Gaius Julius), 100 v.C.-44 v.C.
César, Gai Juli, 100 av-44 ap
Caesar, Gaius Julius
Caesar, Gaius Iulius (100-44 p.n.e.)
Caesar, Gajus Julius, 100-44 f.Kr.
César, Cayo Julio
Caesar, Gaius Iulius, 100-44 a.C.
Juliusz Cezar

ID VIAF: 28625178 (Personne)
permalien: http://viaf.org/viaf/28625178

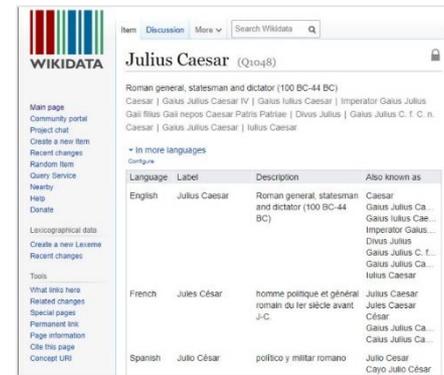
ORCID



HAL



WIKIDATA



WIKIDATA

Item Discussion More Search Wikidata

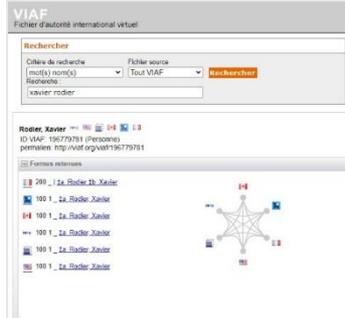
Julius Caesar (Q11048)

Roman general, statesman and dictator (100 BC-44 BC)
Caesar | Gaius Iulius Caesar IV | Gaius Iulius Caesar | Imperator Gaius Julius Gaii filius Gaii nepos Caesar Patris Patrisae | Divus Julius | Gaius Julius C. f. C. n. Caesar | Gaius Iulius Caesar | Iulius Caesar

in more languages

| Language | Label | Description | Also known as |
|----------|---------------|---|--|
| English | Julius Caesar | Roman general, statesman and dictator (100 BC-44 BC) | Caesar Gaius Julius Ca... Gaius Iulius Cae... Imperator Gaius... Divus Julius Gaius Julius C. f... Gaius Julius Ca... Iulius Caesar |
| French | Jules César | homme politique et général romain ou ter siècle avant J.-C. | Julius Caesar Jules Caesar César Gaius Julius Ca... Caïus Julius Ca... Iulius César |
| Spanish | Julio César | político y militar romano | Julio Cesar Cayo Julio César |

Xavier Rodier



VIAF
Fichier d'autorité international virtuel

Rechercher

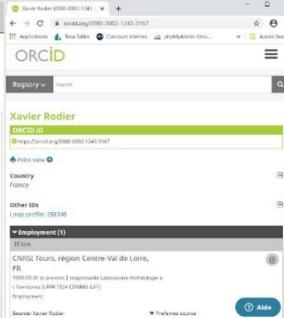
Critère de recherche
mot(s) nom(s) Tout VIAF Rechercher

Recherche
xavier rodier

Rodier, Xavier
ID VIAF: 190779781 (Personne)
permalien: http://viaf.org/viaf/190779781

Formes relatives

- 290 1_ Le Rodier St. Xavier
- 100 1_ Le Rodier Xavier



ORCID

Registry v1

Xavier Rodier

ORCID ID
http://orcid.org/0000-0002-1243-3167

Photo view

country
France

Other IDs
Léonard Rodier: 105745

Employment (1)
CNRS
CNRS Tours, Région Centre-Val de Loire, FR
PROFESOR DE GRADUAT (Emploi(s) Laboratoire Archéologie et Préhistoire)
1 Terrasses (BARR 1024 CHARENTAIS)
Préhistoire

Overview Xavier Rodier



CCSD

MOTS CLÉS
Savoir, Tourisme, Analyse spatiale, Tours, Modélisation, Parcours, Temps, Système d'information géographique, SIG, SIG, Archéologie urbaine, archéologie, Modélisation, Loue, Archéologie, Spatial systems, Commerce, Urban Evolution

COLLABORATEURS

- Laure Badry
- René Gassat
- Elisabeth Zaluska
- Clément Lapeere
- Laure Rodier
- Olivier Millet
- Philippe Hoc
- Sébastien Lefebvre
- Rachel Obez

Xavier Rodier

Nombre de téléchargements: 116



Le nom et le concept

- Problème de l'ambiguïté du nom : Tours, Saint-Martin...
- Problème des traductions : manuscrit, manuscript, manuskript...
- Problème de localisation : coordonnées géographiques, nom de lieu...

**→ Solution : se référer au concept,
pas à une chaîne de caractères dépendant d'une langue**

Lever les ambiguïtés, exemple 1

Corpus 1
« **sanctus Martinus** »
Référence au personnage

Corpus 2
« **saint Martin** »
Référence au personnage

Corpus 3
« **Saint-Martin** »
Référence à un lieu

Comment ne pas confondre deux chaînes de caractères identiques qui font référence à des concepts différents (un personnage / un lieu) ?

Comment associer deux chaînes de caractères différentes qui font référence à un concept identique (différentes appellations pour un même personnage) ?

➔ Utiliser l'identifiant correspondant au concept, indépendant de la langue :

- Personnage : <http://viaf.org/viaf/73825464>
- Lieu : <https://www.geonames.org/3578421/>

Lever les ambiguïtés, exemple 2

« **Vienne** » ?

- Département ?
- Rivière affluente de la Loire ?
- Rivière affluente de la Saône (Seine-Maritime) ?
- Ville d'Isère ?
- Ville autrichienne ?

→ Lever les ambiguïtés :

- Département : <http://www.geonames.org/2969280>
- Rivière affluente de la Loire : <http://www.geonames.org/2969283>
- Rivière affluente de la Saône (Seine-Maritime) : [https://fr.wikipedia.org/wiki/Vienne_\(Seine-Maritime\)](https://fr.wikipedia.org/wiki/Vienne_(Seine-Maritime))
- Ville d'Isère : <http://www.geonames.org/2969284>
- Ville autrichienne : <http://www.geonames.org/2761369>

En pratique...

- Stocker toujours l'identifiant pérenne non ambigu du concept
- Stocker éventuellement l'étiquette (label) associée au concept dans la langue de publication du corpus (plus pratique pour l'utilisateur qu'un identifiant...)
ou
- ✓ Exploiter les services Web pour récupérer dynamiquement les informations associées au concept (étiquette dans une langue donnée, coordonnées géographiques, synonymes associés, terme plus générique dans le thésaurus, etc.)

Quand tout le monde utilise les mêmes référentiels pour décrire les données...

- les données ne sont plus ambiguës ;
- les données sont comparables ;
- les concepts communs forment des ponts entre plusieurs jeux de données.

➔ C'est le premier pas vers l'interopérabilité

Reusable / Réutilisable

R1 Les (méta)données sont richement décrites avec attributs précis et pertinents

R1.1 Les (méta)données sont publiées avec une licence d'utilisation des données claire et accessible

R1.2 Les (méta)données sont associées à une provenance détaillée

R1.3 Les (méta)données répondent à des normes communautaires pertinentes pour le domaine

- (Meta)data are richly described with a plurality of accurate and relevant attributes
 - (Meta)data are released with clear and accessible data usage license
 - (Meta)data are associated with detailed provenance
 - (Meta)data meet domain-relevant community standards

Licences *Creative Commons*

<https://creativecommons.org/licenses/?lang=fr-FR>

- **ATTRIBUTION**

CC BY



- **ATTRIBUTION / PAS DE MODIFICATION**

CC BY-ND



- **ATTRIBUTION / PAS D'UTILISATION COMMERCIALE / PAS DE MODIFICATION**

CC BY-NC-ND



- **ATTRIBUTION / PAS D'UTILISATION COMMERCIALE**

CC BY-NC



- **ATTRIBUTION / PAS D'UTILISATION COMMERCIALE / PARTAGE DANS LES MÊMES CONDITIONS**

CC BY-NC-SA



- **ATTRIBUTION / PARTAGE DANS LES MÊMES CONDITIONS**

CC BY-SA



Licence ouverte Etalab



LICENCE OUVERTE
OPEN LICENCE

<https://www.etalab.gouv.fr/licence-ouverte-open-licence/>

FAIRisation => Partage

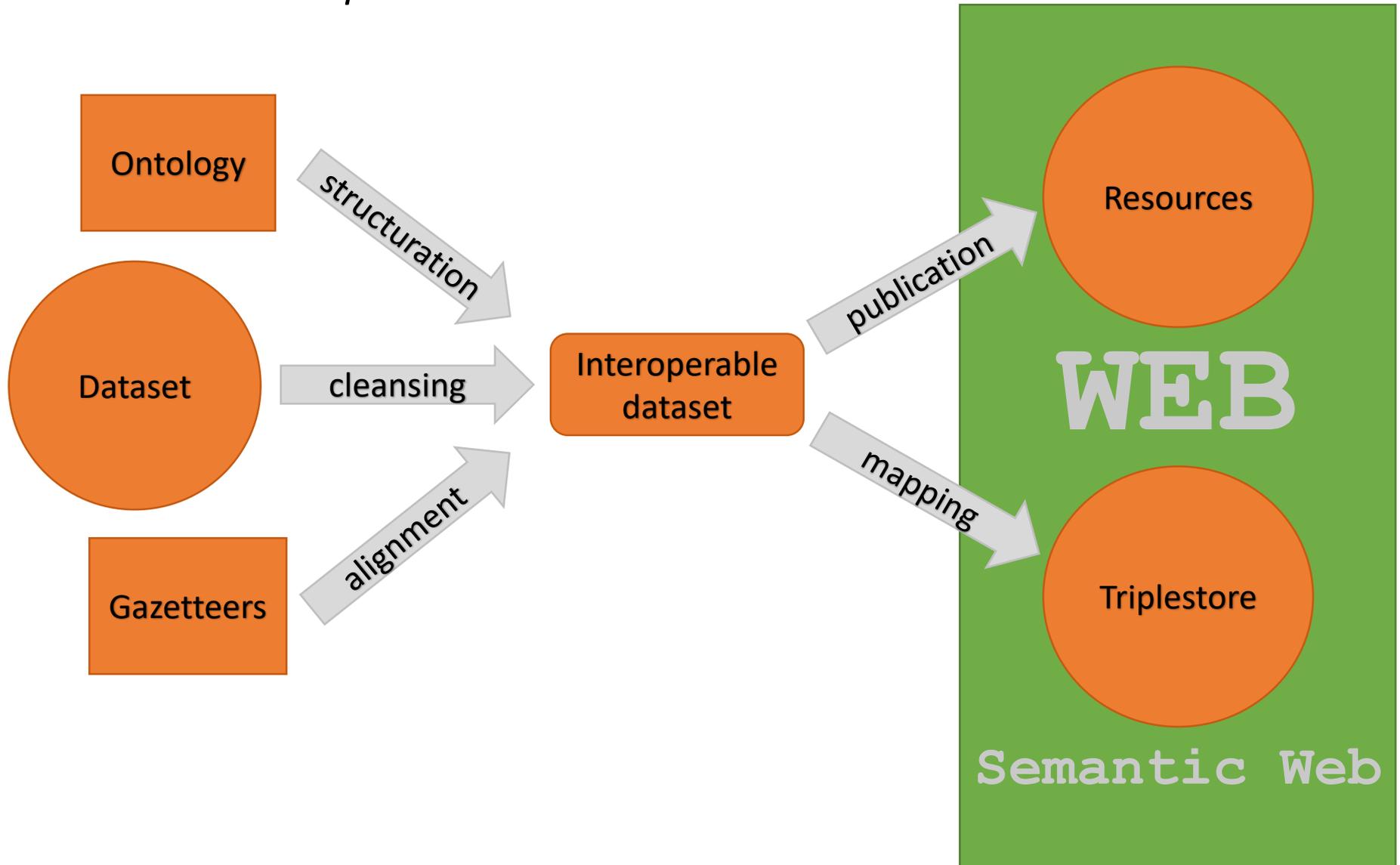
Exprimer les données dans une ou plusieurs ontologies de domaines

Ouvrir, partager ses données

- Créer et devenir un entrepôt de référence du domaine (Certification)
- Exposer les données dans un entrepôt (Connaître l'écosystème du domaine d'application / les Entrepôts de référence du domaine)
 - Par moissonnage (OAI-PMH notices format XML)
 - Par agrégation/import (transformation RDF)
- Les enrichir avec
 - des métadonnées issues de standards
 - des concepts issus de thesaurus interopérables
- Renseigner les provenances, appliquer des licences (meilleure réutilisabilité)
- Utiliser des identifiants pérennes

=> Science reproductible

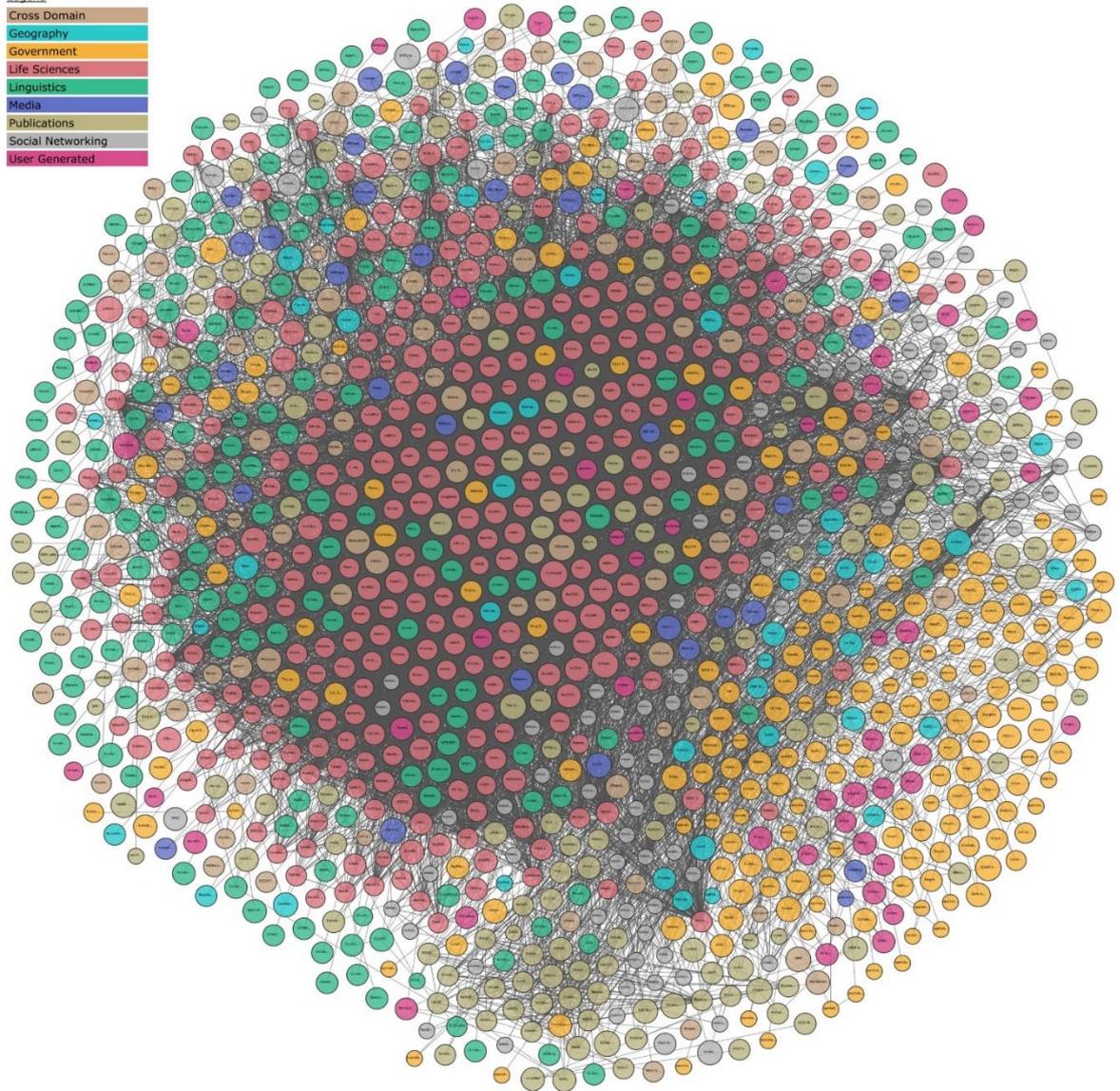
Workflow pour porter les données sur le *Linked Open Data*



Le *Linked Open Data*

Legend

| |
|-------------------|
| Cross Domain |
| Geography |
| Government |
| Life Sciences |
| Linguistics |
| Media |
| Publications |
| Social Networking |
| User Generated |



Contexte de la Science Ouverte

Science Ouverte

1^{er} Plan National pour la Science Ouverte (2018 – 2021)

2^{ème} PNSO (2021 – 2024)

axe 1 : Généraliser l'accès ouvert aux publications

Science Ouverte

1^{er} Plan National pour la Science Ouverte (2018 – 2021)

2^{ème} PNSO (2021 – 2024)

axe 2 : Structurer, partager et ouvrir les données de la recherche

- Création de **Recherche Data Gouv** (inauguration le 8 juillet 2022)
 - service de dépôt et de diffusion des données (entrepôt)
 - catalogue des données de la recherche française (signalement des données)
 - accompagnement des chercheurs (certification des dépôts, mise en œuvre des PGD, transposition de la directive européenne sur le droit d'auteur en matière de fouille de texte et de données)
- Mettre en œuvre l'obligation de diffusion des données de la recherche financées sur fonds publics
- Promouvoir l'adoption de la politique de données sur l'ensemble du cycle de vie (FAIRisation)

Science Ouverte

1^{er} Plan National pour la Science Ouverte (2018 – 2021)

axe 3: s'inscrire dans une dynamique durable, européenne et internationale

Contribuer à la structuration de l'European Open Science Cloud et (CNRS) s'inscrire dans la mise en place des services de données qui s'y installent, participer à GO FAIR (MESRI)

2^{ème} PNSO (2021 – 2024)

axe 3 : Ouvrir et promouvoir les codes sources produits par la recherche
Soutenir Software Heritage, proposer la standardisation du « Software Heritage Identifier » ou « SWHID » qui complète les DOI

axe 4 : Transformer les pratiques pour faire de la Science ouverte le principe par défaut

Reconnaitre la Science ouverte dans le système d'évaluation

Science Ouverte

Mise en œuvre pratique du Plan

- avec le Comité pour la science ouverte (MESRI)

<https://www.ouvrirlascience.fr/accueil/>

Collèges et groupes d'expertises (FNSO)

- avec la DDOR (CNRS)

<https://www.science-ouverte.cnrs.fr/>

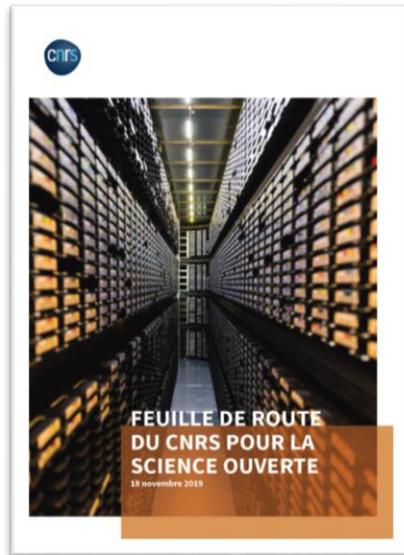
Feuille de route (novembre 2019)

https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/Plaqueette_Science-Ouverte_18112019.pdf

Plan données de la recherche du CNRS (novembre 2020)

https://www.science-ouverte.cnrs.fr/wp-content/uploads/2021/01/Plaqueette-Plan-Donnees-Recherche-CNRS_nov2020.pdf

Science Ouverte



FAIR
DATA



recherche.[data.gouv.fr](https://www.data.gouv.fr)

OPEN
SCIENCE

Science Ouverte

Mise en œuvre pratique du Plan (CNRS)

Inciter la FAIRisation le plus en amont possible

Déposer ses données dans un entrepôt en accès libre

Aller vers la certification de ces entrepôts

Encourager la réutilisation plutôt que la création

Comment , dans quels entrepôts diffuser les données ?

<https://www.ccsd.cnrs.fr/2016/11/comment-diffuser-mes-donnees-de-recherche-1/>

<https://www.ccsd.cnrs.fr/2016/11/comment-diffuser-mes-donnees-de-recherche-2/>

EOSC, ARIADNE+ (archéologie), Conservatoire des données 3D, Nakala, HAL, DRYAD

SHS : Huma-Num pour accompagner la communauté avec ses consoritum

Contexte : La Science Ouverte (OpenScience)

- Objectif : partager, réutiliser les données, économies d'échelle, interrogations plus efficace

- Beaucoup de réseaux européens



- Tout domaine



- Sciences-Humaines



- Sciences patrimoniales



- Archéologie

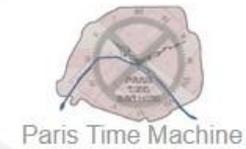
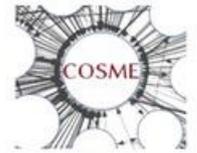


ACCOMPAGNER LES COMMUNAUTÉS DE RECHERCHE



Huma-Num^{IR*}

- Les consortium labellisés



Bonnes pratiques

Bonnes pratiques

Cycle de vie
des données de la recherche

Questions à se poser
pour tout nouveau projet



Bonnes pratiques – Avant

Explorer les ressources disponibles en lien avec mon projet

<https://marketplace.sshopencloud.eu/>



Bonnes pratiques – Avant

Explorer les ressources disponibles en lien avec mon projet

The screenshot shows the SSH Open Marketplace website. At the top left is the logo for SSH Open Marketplace. The navigation menu includes: Tools & services, Training materials, Publications, Datasets, Workflows, Browse, Contribute, and About. On the right side of the header, there are links for 'Report an issue' and 'Sign in'. The main heading is 'Social Sciences & Humanities Open Marketplace'. Below this is a sub-heading: 'Discover new and contextualised resources for your research in Social Sciences and Humanities: tools, services, training materials, workflows and datasets.' followed by a 'Read more...' link. A search bar is present with a dropdown menu for 'All categories' and a 'Search' button. The background features an illustration of people working together in a cityscape with landmarks like the Eiffel Tower and the London Eye. Below the search bar, there are two sections: 'Browse' and 'Last updated'. The 'Browse' section has a 'Browse by Activity' link and a 'See all' link. The 'Last updated' section has a 'See what's new' link and a 'See all' link. The 'Browse by Activity' section lists various activities with their respective counts: Analyzing (644), Data Visualization (371), Visual Analysis (309), Content Analysis (256), Discovering (207), Annotating (173), Capturing (171), Enriching (148), Collaborating (144), Disseminating (142), Publishing (129), Creating (124), Gathering (121), Organizing (111), Sharing (95), Editing (90), Storing (84), Network Analysis (81), Data Cleansing (61), and Writing (59). The 'Last updated' section features a blue icon and the text 'SPPAS', with sub-headers for 'Activities' (Annotating, Audio Annotation) and 'Keywords' (Annotating). A description below reads: 'SPPAS - the automatic annotation and analyses of'.

Bonnes pratiques – Avant

S'informer au niveau local / national pour l'environnement du projet

1. Y-a-t-il des infrastructures pour la conservation des données, leur archivage à long terme ?
2. Y-a-t-il des outils à disposition pour gérer (traiter/enrichir) les données le temps du projet ?
3. Comment les principes FAIR sont-ils implémentés sur ces outils ?

Bonnes pratiques – Avant

Préparer la rédaction du plan de gestion des données (DMP/PGD)

- Utiliser les outils de l'INIST (DMP OPIDOR)
- Pointer tous les éléments essentiels à l'implémentation des principes FAIR sur les données du projet

<https://datapartage.inrae.fr/Produire-des-donnees-FAIR>

Bonnes pratiques – Pendant

Bonnes pratiques – Pendant

Les données

Quelles données seront produites ?

Où stocker les données pendant le projet ? Est-ce un endroit sécurisé ?

Bonnes pratiques – Pendant

Le traitement des données

Quel type de processus utiliser ?

Avec quel(s) logiciel(s) ?

Cela implique quel type de format ? est-il ouvert ? si non, est-ce possible de convertir les fichiers ? dans quels formats ouverts ?

Bonnes pratiques – Après

Bonnes pratiques – Après

Archiver les données

Quelles données archiver ? Sur quel entrepôt ?

Bonnes pratiques – Après

Partager, Publier, Diffuser les données

Lesquelles et comment les disséminer ? (moissonnage ? extractions/mapping)

Sur quels portails ?

« Comment diffuser mes données de recherche »

<https://openarchiv.hypotheses.org/3961>

« Guide pratique – Ouvrir la science »

https://www.ouvrirlascience.fr/wp-content/uploads/2021/10/SO_21-10-14-IMPR-FR.pdf

Bonnes pratiques



<https://www.go-fair.org/>

Application des principes FAIR

Auto-évaluation :

https://dorum.fr/enjeux-benefices/outil-fair-aware_10_13143_ege0-nw05/

<https://dorum.fr/appli-fair-aware-pleine-page-vf/>

Sources d'information:

https://www.go-fair.org/wp-content/uploads/2022/01/FAIRPrinciples_overview.pdf

en français : Recommandation pour FAIRiser vos données (PARTHENOS)

<https://zenodo.org/record/3463521#.YqmMvOzP2Uk>

Sources d'information

CAT OPIDOR : <https://cat.opidor.fr>

Pour s'y retrouver dans les services existants, le catalogue des services français qui contribuent à la gestion, au partage et à la réutilisation des données de la recherche.

DMP OPIDOR : <https://dmp.opidor.fr/>

Pour aider à la constitution d'un Plan de Gestion des Données (Data Management Plan)

DORANUM : <https://doranum.fr/>

Des ressources pour accompagner la communauté scientifique dans la gestion et le partage de leurs données

Ressource du site Ouvrir la science : <https://www.ouvrirlascience.fr/category/ressources/>

...

Les plateformes existantes

Plateformes pour les données 3D

- Documentation des données
 - aLTAG3D prépare les données
- Archivage des données :
 - le CINES est la plateforme pour l'archivage
- Diffusion et exposition des données 3D
 - Le Conservatoire Nationale des Données 3D
 - Nakala (2D friendly) / Isidore
 - OpenArcheo pour les données archéologiques
 - ...

Quelques bonnes adresses



Conservatoire

- CND3D <https://3d.humanities.science>



aLTAG3D

- aLTAG3D <https://altag3d.huma-num.fr>
- aLTAG3D documentation: <https://altag3d.huma-num.fr/docs>



Consortium 3D SHS

- Blog : <https://shs3d.hypotheses.org/>
- Twitter: https://twitter.com/consortium_3d?lang=fr

Les recommandations du Consortium 3D SHS

<https://hal.archives-ouvertes.fr/hal-01683842/document>