

# Données relationnelles hétérogènes en SHS

Anne Garcia Fernandez

Elisabeth Greslou

Univ. Grenoble Alpes, CNRS, Litt&Arts, 38000 Grenoble, France

*avec l'aimable participation de Camille Desiles*



Les SHS





texte

bases de données



objets physiques



image



données bibliographiques

# Et pourtant...

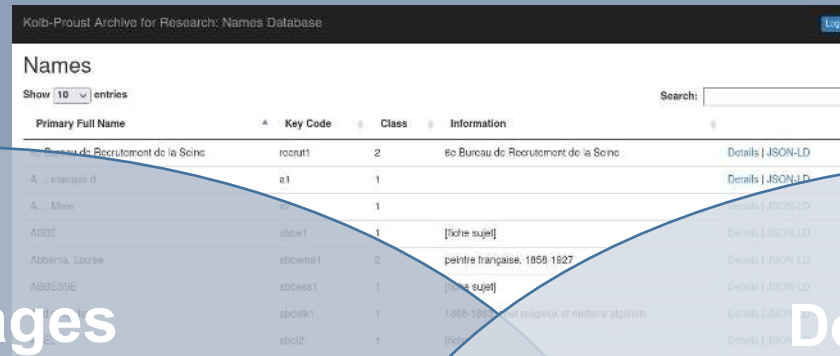
Hétérogénéité des données

Transcription  
Alignement  
Extraction  
Balisage  
Zonage  
...

et des traitements

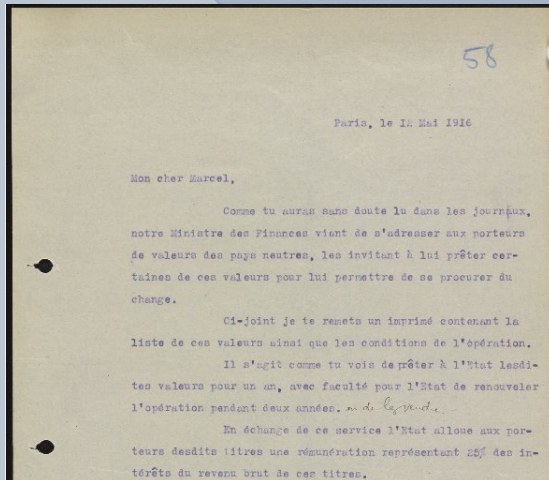
# Exemple 1 : Le projet Corr-Proust *Marcel Proust*

Une base de donnée externe



Primary Full Name	Key Code	Class	Information	
Bureau de Recrutement de la Seine	recrut1	2	ec Bureau de Recrutement de la Seine	Details   JSON-LD
A. ... marquis d	r1	1		Details   JSON-LD
A. ... Mme		1		Details   JSON-LD
ABBE	abbe1	1	[sûre sujet]	Details   JSON-LD
Abbeix, Louise	abbeix1	2	peintre française, 1858-1927	Details   JSON-LD
ABESSE	abesse1	1	[sûre sujet]	Details   JSON-LD
	abdok1	1	1806-1850, religieux et militaire algérien	Details   JSON-LD
	abd2	1	[sûre sujet]	Details   JSON-LD

Des images



Des transcriptions

Paris, le 12 Mai 1916

Mon cher **Marcel**,

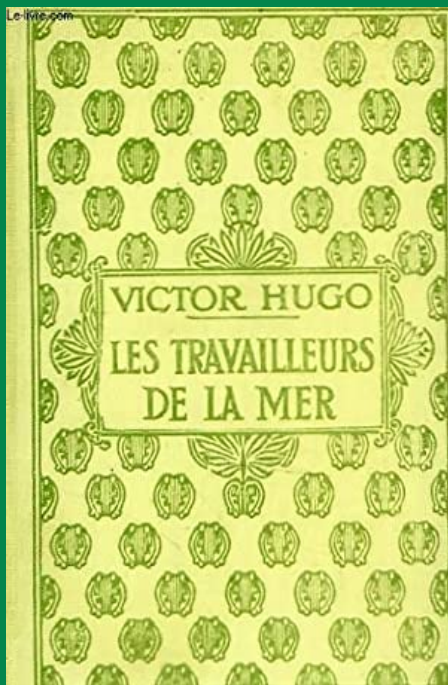
Comme tu auras sans doute lu dans les journaux, notre **Ministre des Finances** vient de s'adresser aux porteurs de valeurs des pays neutres, les invitant à lui prêter certaines de ces valeurs pour lui permettre de se procurer du change<sup>1</sup>.

Ci-joint je te remets un imprimé contenant la liste de ces valeurs ainsi que les conditions de

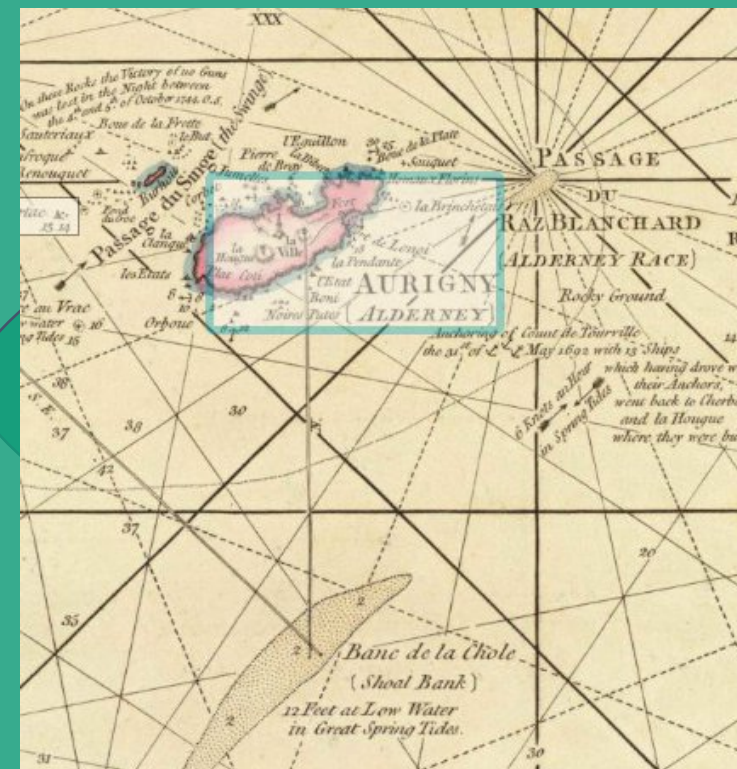
site

# Exemple 2 : Carto-Hugo

Édition textuelle



Images



Zonage

# Exemple 3 : Données bibliographiques



Sévigné



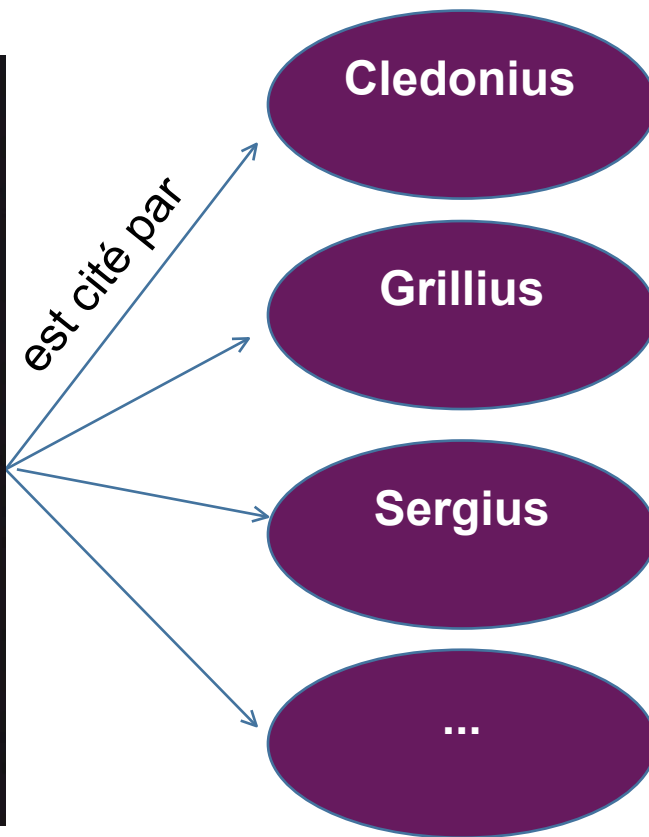
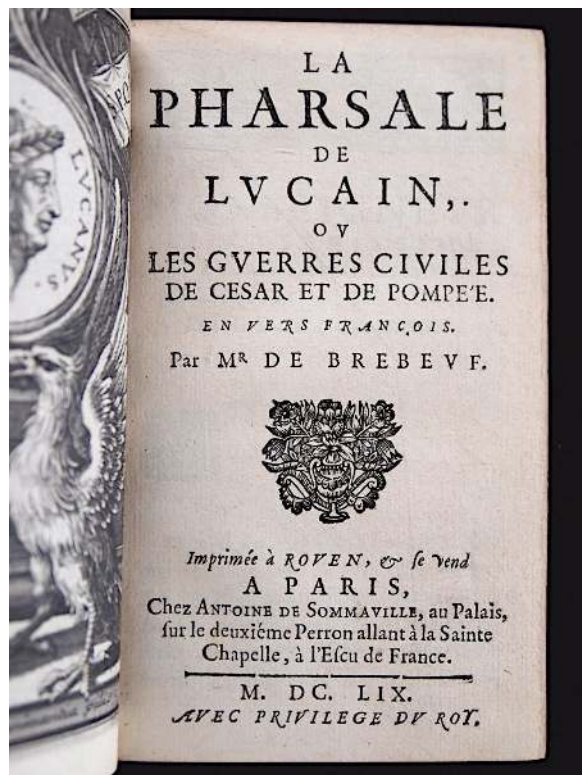
L'Observatoire des  
littératures  
francophones du Sud



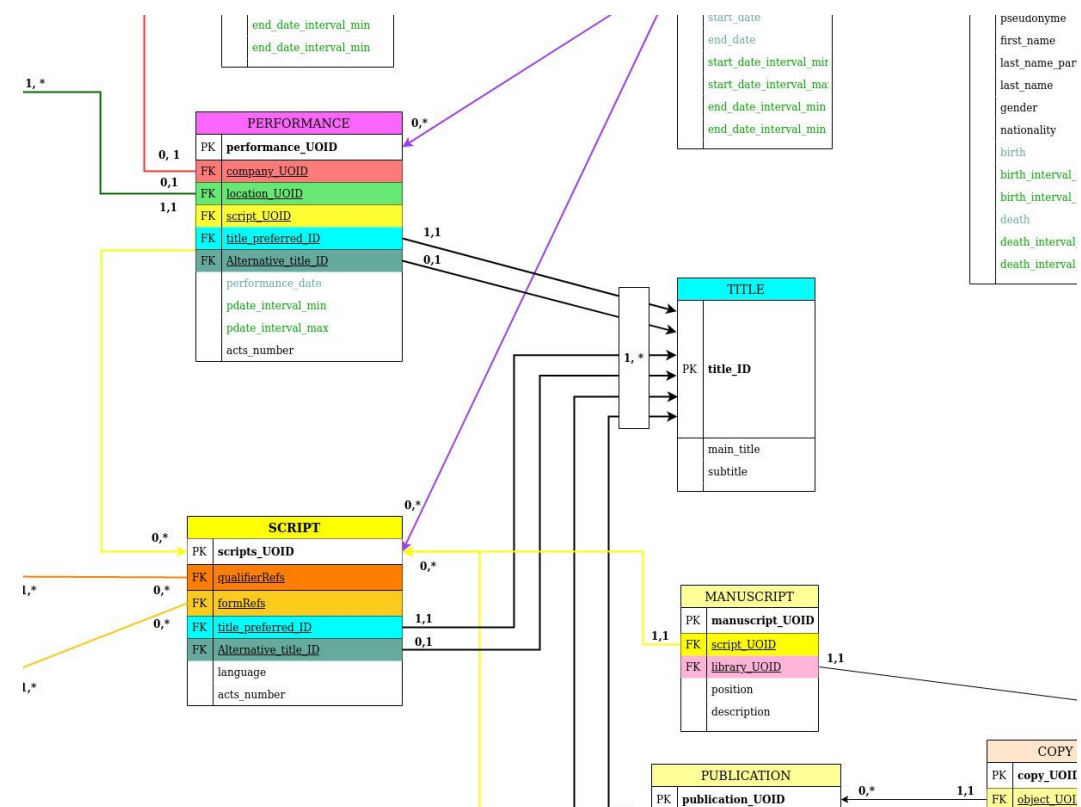
Données  
bibliographiques

# Exemple 4 : Données liées

A QUO



CESAR



# Les effets du temps...

## quelques exemples



CESAR a été fondé en 2001 par [Barry Russell \(1943-2003\)](#) à Oxford Brookes University avec le soutien du British Arts and Humanities Research Council (AHRC). La base de données formant le cœur de CESAR avait été construite en collaboration avec [David Trott \(1940-2005\)](#), alors professeur à l'Université de Toronto, et Jeffrey Ravel, professeur au Massachusetts Institute of Technology, qui avaient apporté leur propre documentation. En 2004, Oxford Brookes University a obtenu une nouvelle subvention de l'AHRC pour créer une banque d'images. Le site a été alors dirigé par Mark Bannister, puis par Sabine Chaouche. Le site qui est un outil collaboratif ouvert à tous les spécialistes et amateurs de théâtre, a été régulièrement alimenté par des chercheurs français et internationaux, notamment Georges Forestier et Anastasia Sakhnovskaia Pankeeva.

**Grâce à un accord entre Oxford Brookes University et l'Université Grenoble Alpes, CESAR est accueilli depuis 2016 par le laboratoire Litt&Arts (UMR CNRS 5316). Sa rénovation technique, soutenue par le CNRS, par l'Institut universitaire de France et par l>IDEX Université Grenoble Alpes, est en cours. Le site et les données sont hébergés par la TGIR Huma-Num.** Dans la mesure où nous travaillons actuellement à une nouvelle version du site (publiée à l'horizon 2020), celle-ci n'est plus maintenue mais reste parfaitement utilisable. Malgré tous nos efforts, quelques fonctionnalités, notamment dans la partie « Livres », peuvent être indisponibles. L'accès aux comptes personnels et à l'espace collaboratif est également temporairement désactivé, mais sera rétabli dans la nouvelle version. Nous vous remercions de votre patience et de votre compréhension.

Tous les contributeurs, anciens, présents et à venir, sont invités à participer aux chantiers de réflexion prochainement ouverts autour de la mise à jour bibliographique, de l'élargissement chronologique de la documentation (XVe-XVIIIe siècle) et d'appels à projets visant au développement thématique et documentaire du site. Une journée d'étude sera organisée en 2020, dans la lignée des précédents [colloques CESAR](#), afin d'envisager collectivement les nouvelles perspectives de ce projet.

Votre soutien dans cette phase décisive permettra à CESAR de rester un outil de référence internationalement reconnu pour l'histoire des arts du spectacle.

Janvier 2017

## Bienvenue sur CESAR

**CESAR est une ressource en ligne dynamique**, mise au service de tous ceux qui s'intéressent au théâtre français des XVIIe et XVIIIe siècles — étudiants, chercheurs, spécialistes du sujet, ou simples amateurs de théâtre.

Le site offre deux principales sources d'information liées entre elles — une base de données et une banque d'images — mais aussi des ressources supplémentaires (version en ligne de l'intégralité des anthologies des Frères Parfaict et de Lérís, sélection de comptes-rendus de l'époque, rapports de police et textes théoriques). Tous ces documents peuvent être consultés à partir de la base de données ou de la banque d'images.

[Accéder à la base de données](#)

[Accéder à la banque d'images](#)

CESAR est fondé sur la coopération active de ses membres. L'accès en est gratuit. Chercheurs et spécialistes travaillant sur le Théâtre de l'Ancien Régime ou durant la période révolutionnaire sont invités à faire part de leurs suggestions afin d'assurer la fiabilité et la clarté de la base de données. Tout renseignement, toute annotation ou correction sont les bienvenus, de même que toute proposition de collaboration pour développer le site.

[À propos de CESAR](#)  
[Sources et Collaborateurs](#)  
[Liste de diffusion](#)

## Welcome to CESAR

**CESAR is a dynamic resource** at the service of all those with an interest in the French theatre of the seventeenth and eighteenth centuries — students, researchers, scholars, enthusiasts.

It offers two major interlinked facilities — a database and an imagebank — but also supplementary materials (complete online versions of key compendia such as Parfaict and Lérís, a corpus of contemporary reviews, police reports and treatises), all of which are accessible from either the database or the imagebank.

[Enter the database](#)

[Enter the imagebank](#)

CESAR is conceived as a genuinely collective enterprise. Its extensive contents are freely available and it is hoped that scholars working on any aspect of Ancien Régime and Revolutionary theatre will help to make this resource even more comprehensive and as reliable as possible by contributing data, annotations and corrections and by offering support in developing the site.

[About CESAR](#)  
[Sources and Contributors](#)  
[Newsletter](#)



# Les manuscrits de Stendhal

3667 pages  
ont été transcrites  
et sont à la disposition du  
public.

Ce site vous permet d'accéder aux manuscrits de Stendhal conservés à la Bibliothèque municipale de Grenoble :  
vues numérisées des pages, transcriptions et descriptions.

Un projet né de la collaboration entre la Ville de Grenoble,  
l'Université Stendhal - Grenoble 3 et la MSH-Alpes.

## Zoom sur

### Arenc, Arrhain, Allauch, Avrain ?

Jean-Jacques Labia a enquêté sur ce toponyme  
qui désigne un lieu fréquenté par Stendhal.

[Pour en savoir plus...](#)

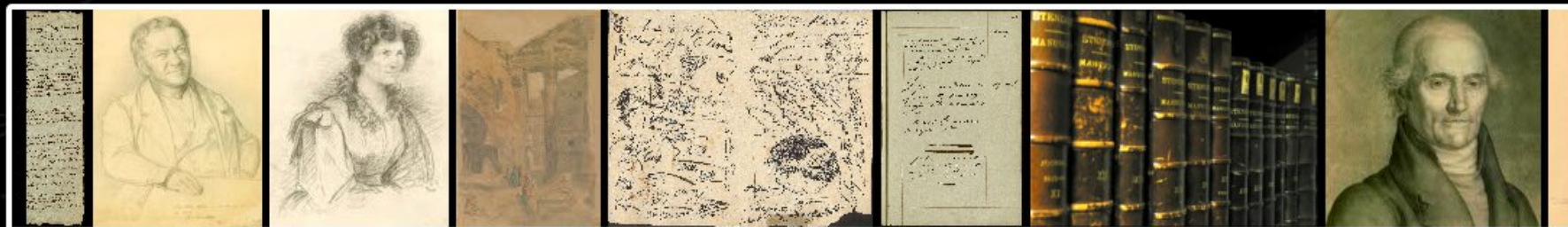
### Stendhal et le latin

Voir sur ce sujet la note de Cécile Meynard

[Pour en savoir plus...](#)

### Comment le "corbeau" est devenu Falconi

Pour le savoir, lire la note d'Hélène de Jacquelot



## Accédez aux manuscrits

# Projet Réticence



[Accueil](#) [Le projet](#) [L'inventaire](#) [Exploration des brouillons](#) [Expérimentations](#) [Créations](#) [À propos](#)

## Explorations des brouillons de Jean-Philippe Toussaint

Mettre en ligne les brouillons pour permettre à chacun d'accéder aux archives de la création littéraire, de les décrypter, et de s'en emparer pour de nouvelles recherches ou de nouvelles créations : tel est l'objectif du projet élaboré à partir des brouillons du quatrième roman de [Jean-Philippe Toussaint](#), *La Réticence*, publié aux Éditions de Minuit en 1991.



Dessin figurant sur [le facsimilé LRT\\_02\\_01\\_00058](#).

# Des solutions pour la pérennité

# La mise sous cloche

L'exemple des Manuscrits de Stendhal



# La mise sous cloche

L'exemple des Manuscrits de Stendhal



ou plutôt



Merci :



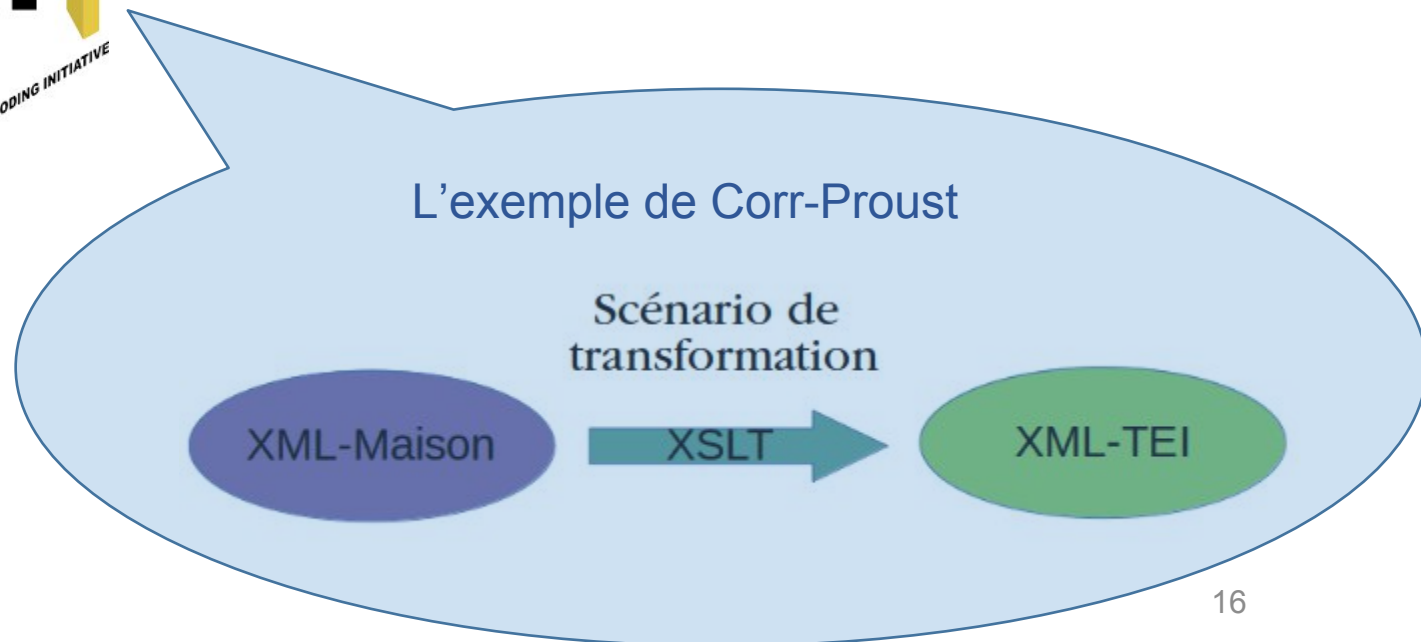
# Suivre les standards

Données, métadonnées, logiciels, valeurs de métadonnées, etc.



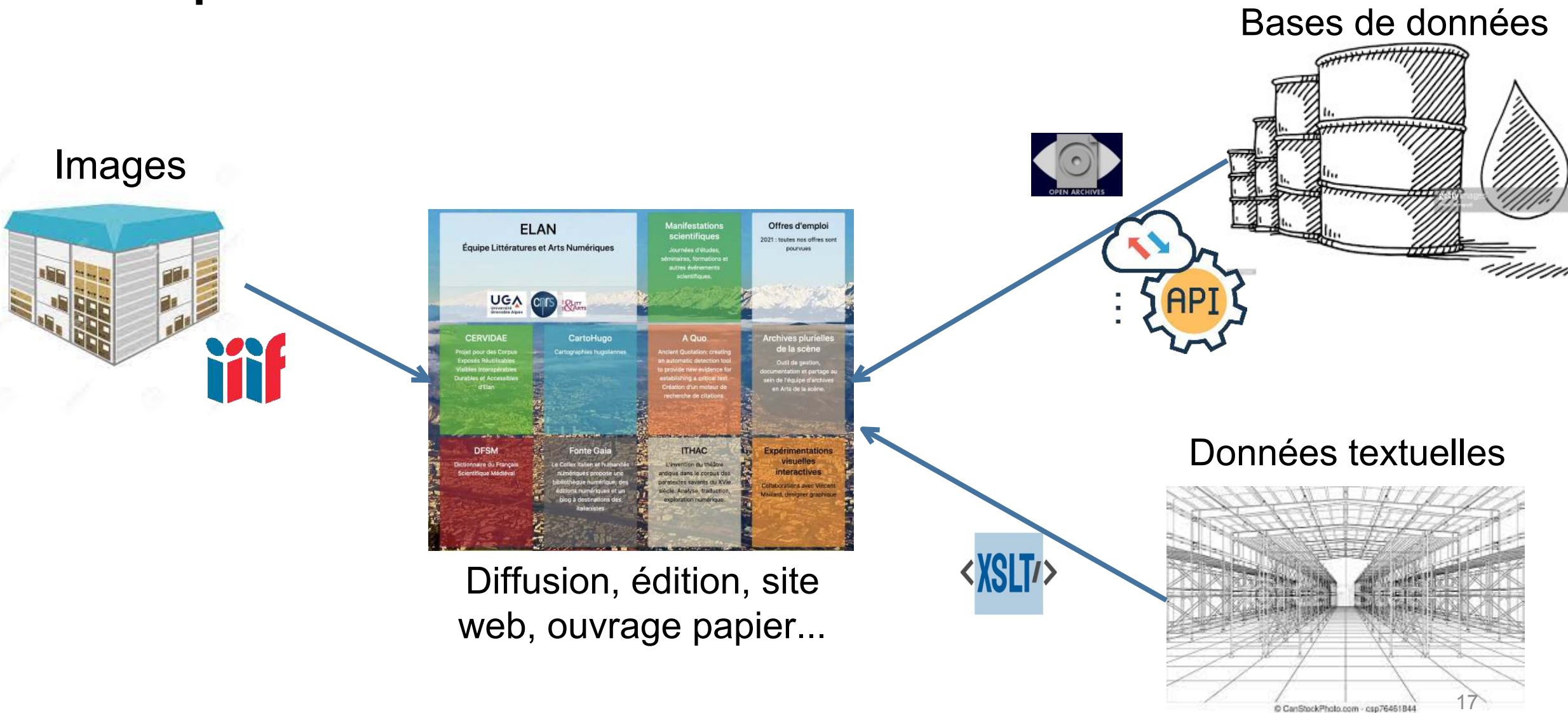
# Suivre les standards

Données, métadonnées, logiciels, valeurs de métadonnées, etc.

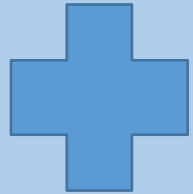




# Séparer les données des interfaces



# Projet : Carto Hugo



TEI

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet href="..." type="text/css"?>
3 <svg xmlns="http://www.w3.org/2000/svg" xmlns:cc="http://creativecommons.org/ns#" x
4 <image xlink:href="..." height="11730" width="7600" />
5 <g class="location location-hidden" data-location="Ancrese">
6 <title>Ancrese (L')</title>
7 <rect width="108" height="47" x="1267" y="4329" />
8 </g>
9 <g class="location location-hidden" data-location="Bretagne">
10 <title>Bretagne</title>
11 <rect width="0" height="0" />
12 </g>
13 <g class="location location-hidden" data-location="Calvados">
14 <title>Calvados</title>
15 <rect width="0" height="0" />
16 </g>
17 <g class="location location-hidden" data-location="Carteret">
18 <title>Carteret</title>
19 <rect width="0" height="0" />
20 </g>
21 </svg>
```



Victor Hugo, Les Travailleurs de la mer

## V LES RISQUES DE MER

L'overfall, lisez : casse-cou, est partout sur la côte ouest de Guernesey. Les vagues l'ont savamment déchiquetée. La nuit, sur la pointe des rochers suspects, des clartés invraisemblables, aperçues, dit-on, et affirmées par des rôdeurs de mer, avertissent ou trompent. Ces mêmes rôdeurs, hardis et crédules, distinguent sous l'eau l'holothurion des légendes, cette ortie marine et infernale qu'on ne peut toucher sans que la main prenne feu. Telle dénomination locale, Tinttajeu, par exemple (du gallois, Tin-Tagel), indique la présence du diable. Eustache, qui est Wace, le dit dans ses vieux vers :

Dont commença mer à meller ;  
Undes à croïstres et à troubler,  
Noircir il cieus, noircir la nue ;  
Tost fust la mer toute espandue.

Cette Manche est aussi inconnue aujourd'hui au feu temps de Toudrin, d'Imbrofal

Lieu  < >

Carto Hugo - Anna

Lieu Baise Saint-Simpson

Victor Hugo, Les Travailleurs de la mer

ramorso, et à droite la balise de Saint-Simpson qui représente une figure d'homme. Il crut apercevoir une troisième balise. Qu'est-ce que cette balise ? Quand l'eût-on plantée sur ce point ? Quel bas-fond indiquait-elle ? La balise répondait tout de suite à ces interrogations, elle venait ; c'était un mat. L'ordonnement du pêcheur ne dénotait point. Une balise faisait question, un mat bien plus encore. Il n'y avait point de pêche possible. Quand tout le monde rentrait, oueki, un soir-là. Qui ? Pourquoi ? Dix minutes après, le mat s'élevait lentement, et l'on voyait la silhouette du pêcheur d'Orpède. Il ne put raconter la balise. Il se borna à dire : il n'y avait que le bras de deux hommes. C'était donc vraisemblablement un homme seul. Le vent était nord ; cet homme évidemment saignait pour aller prandre le vent au delà de la pointe Fontenelle. I. A.

# Projet : Tacitus On Line



Rechercher dans NAKALÀ...

Tacitus On Line  
ID : 11280/3ab03291 | Publique  
Créé le 16/07/2020

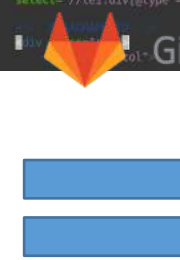
- Tacti et C. Velleii Paterculi scripta quae exstant; recognita, emaculata. Additque commentarii copiosissimi et notae non antea editae Paris e typographia Petri Chevalier, in monte diui Hilarii - III-0153
- Tacti et C. Velleii Paterculi scripta quae exstant; recognita, emaculata. Additque commentarii copiosissimi et notae non antea editae Paris e typographia Petri Chevalier, in monte diui Hilarii - III-0151
- Tacti et C. Velleii Paterculi scripta quae exstant; recognita, emaculata. Additque commentarii copiosissimi et notae non antea editae Paris e typographia Petri Chevalier, in monte diui Hilarii - III-0150



```

1 <?xml version="1.0" encoding="UTF-8" ?>
2
3 <!-- transform: xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
4 xmlns:as="http://www.w3.org/2001/XMLSchema" exclude-result-prefixes="xsl"
5 xmlns:tei="http://www.tei-c.org/ns/1.0" ?-->
6
7 <xsl:method method="html" omit-xml-declaration="yes" indent="no"
8 <xsl:output method="html" indent="no"
9 <xsl:output name="typeMap"
10 <xsl:output select="//tei:list[@xml:id='typology_comments']"
11 </xsl:output>
12 <xsl:attribute name="DEBUG" as="xs:boolean" true />
13
14 <xsl:namespace-alias namespace-prefix="tei" />
15
16 <xsl:variable name="current-liber"
17 <xsl:variable name="current-liber"
18 <xsl:variable select="*" />
19
20 <xsl:attribute href=".../views/html/Liber{subscript}
21 method="html" indent="yes" omit-xml-declaration="no"
22 <xsl:attribute select="//tei:div[@type='liber' and @xml:id='
23
24
25
26
27

```



```

- <div type="paragraph" xml:id="n1.1" ?>
- <span type="sub-paragraph" xml:id="n1.1.1" ?>
  Urbem Romam a principio reges habuere; libertatem et consulatum L.
  Brutus instituit. Dictaturae ad tempus sumebantur; neque decemviralis
  potestas ultra biennium, neque tribunorum militum consularis ius diu
  ualuit. Non Clauae, non Sullae longa dominatio; et Pompei Crassique
  potentia cito in Caesarem, Lepidi atque Antonii arma in Augustum
  cessere, qui cuncta, discordis ciuilibus fessa, nomine principis sub
  imperium accepit.
</span>
- <span xml:id="n1.1.2" type="sub-paragraph" ?>
  Sed ueteris populi Romani prospera uel aduersa claris scriptoribus
  memorata sunt; temporibusque Augusti dicendis non defuere decora
  ingenia, donec gliscente adulatione detererentur. Tiberii Caique et
  Claudii ac Neronis res florentibus ipsis ob metum falsae, postquam
  occiderant recentibus (5)odiis compositae sunt.
</span>
- <span xml:id="n1.1.3" type="sub-paragraph" ?>
  Inde consilium mihi pauca de Augusto et extrema tradere, mox Tiberii
  principatum et cetera, sine ira et studio, quorum causam prociui habeo.
</span>
</div>
- <div type="text" xml:id="n1.2" ?>
  - <span type="sub-paragraph" xml:id="n1.2.1" ?>

```

The screenshot shows the web interface of the Tacitus On Line project. On the left, there is a sidebar with navigation options like 'Menu', 'Commentaire', 'Index', etc. The main content area displays a list of works, with the first entry selected. The selected entry shows the title 'Tacti et C. Velleii Paterculi scripta quae exstant; recognita, emaculata. Additque commentarii copiosissimi et notae non antea editae Paris e typographia Petri Chevalier, in monte diui Hilarii - III-0153' and a detailed description of the work's content and history.

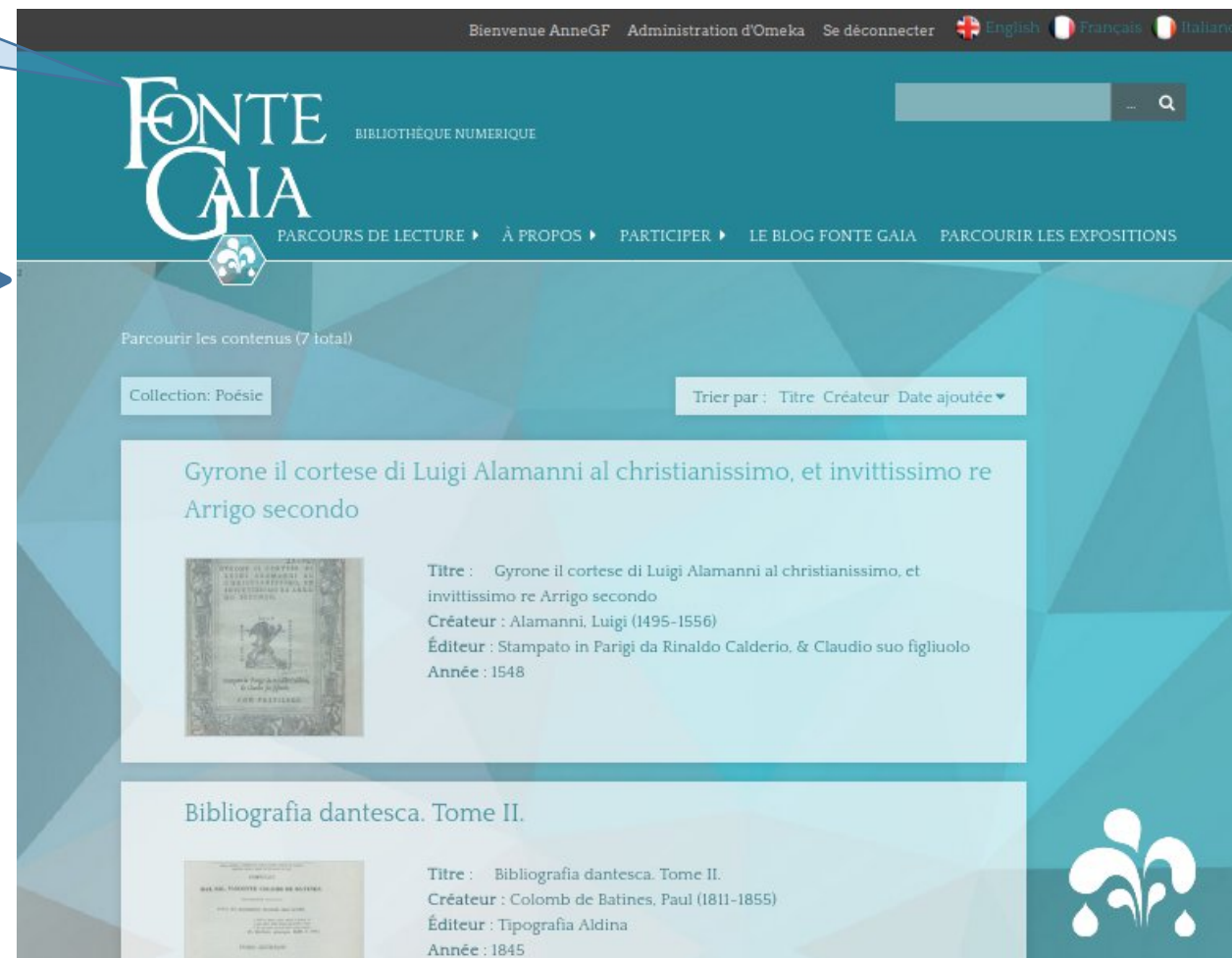


# Ré-utiliser plutôt que re-créeer

L'exemple du projet Fonte Gaia



OAI-PMH

The image is a screenshot of the 'FONTE GAIA BIBLIOTHEQUE NUMERIQUE' website. The header includes navigation links like 'PARCOURS DE LECTURE', 'À PROPOS', 'PARTICIPER', 'LE BLOG FONTE GAIA', and 'PARCOURIR LES EXPOSITIONS'. The main content area shows a list of digital collections. The first item is 'Gyrone il cortese di Luigi Alamanni al christianissimo, et invittissimo re Arrigo secondo' with a small image of the book cover and metadata: 'Titre : Gyrone il cortese di Luigi Alamanni al christianissimo, et invittissimo re Arrigo secondo', 'Créateur : Alamanni, Luigi (1495-1556)', 'Éditeur : Stampato in Parigi da Rinaldo Calderio, & Claudio suo figliuolo', and 'Année : 1548'. The second item is 'Bibliografia dantesca. Tome II.' with metadata: 'Titre : Bibliografia dantesca. Tome II.', 'Créateur : Colomb de Batines, Paul (1811-1855)', 'Éditeur : Tipografia Aldina', and 'Année : 1845'. The website has a teal and blue color scheme.

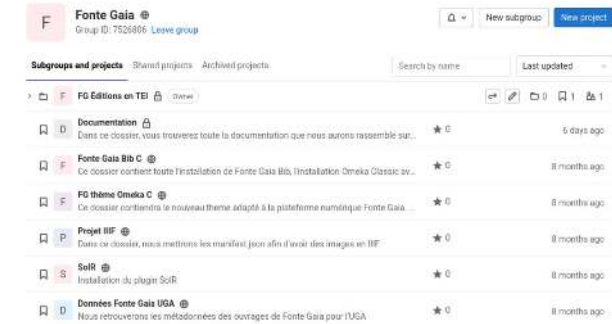
Valable pour les données mais aussi pour les outils, services, processus...

- ex. service de blog Hypothèses.org plutôt que création d'un site web

# Documenter

## Utilisation de Git

- toutes les sources (omeka, plugin, manuels...) sont versionnées



The screenshot shows the website for 'Manuel d'encodage pour la transcription'. The page has a dark red header with navigation links. The main content area is white with a dark red sidebar on the left. The sidebar contains links to various sections like 'Avant-propos', 'Le projet Tacitus On Line', 'Vocabulaire du projet', etc. The main content area has a dark red header with the title 'Manuel d'encodage pour la transcription' and a sub-header 'Avant-propos'. Below this, there are two sections: 'Le projet Tacitus On Line' and 'Avant-propos', each with a list of bullet points.



# Savoir clore...

- Penser à la fin dès le début



- Agir en conséquence

- décrire
- archiver
- transmettre
- ...



# Des solutions émergentes formalisées

FAIR / Linked Open Data / Science ouverte

# Suivre les principes FAIR

Findable   
Accessible   
Interoperable   
Reusable 

## Les principes FAIR

Les chercheurs s'appuient sur les connaissances scientifiques antérieures, notamment sur les résultats publiés dans les articles scientifiques. La reproductibilité des résultats, ainsi que leur croisement, ne sont cependant envisageables qu'avec des données originelles et leurs conditions d'obtention. C'est pourquoi la science ouverte vise à faciliter l'accès aux publications scientifiques et aux données de la recherche. Cette facilitation s'accompagne d'un certain nombre de mesures pour rendre les données scientifiques facilement découvrables, accessibles, interoperables et réutilisables. Ce sont les principes FAIR : Findable, Accessible, Interoperable, Reusable.



URFIST

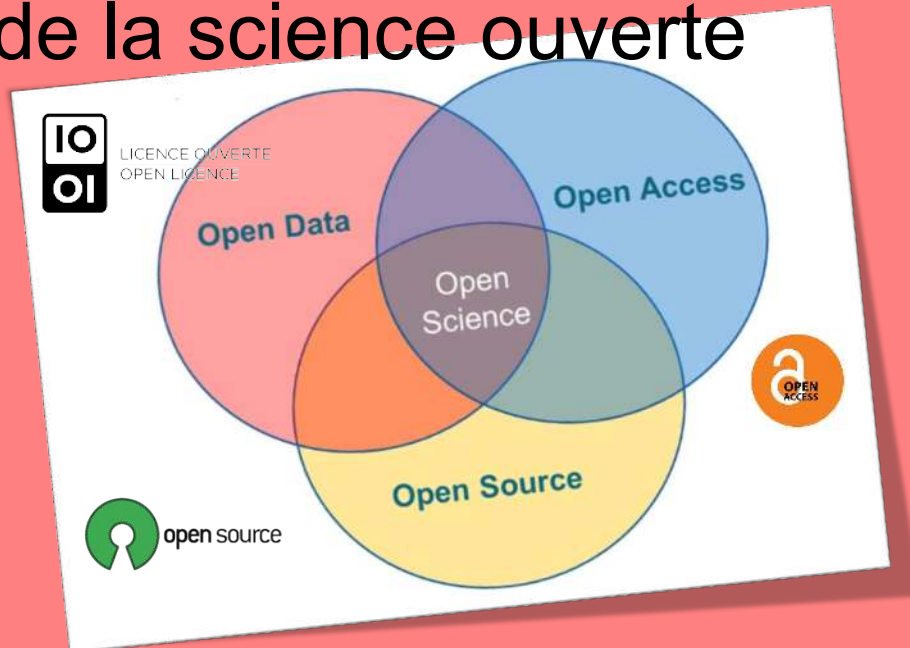
Références ...

# Faire du Linked Open Data



Web sémantique : attention *pas* technique important !

# Faire de la science ouverte



## Faire un plan de gestion de données

- Penser à l'après dès le début



## Publier des Data Papers

- Communiquer, transmettre...



# Conclusion



**Merci de votre attention !**

Rideau