COSME^2 - Complexities: 30 Years Of Research Of Medievalists DH Concerning A Thousand Years Of Medieval Sources

Paul Bertrand, Ariane Pinche, Matthias Gille Levenson, Margot Ferrand

▶ **To cite this version:**

Paul Bertrand, Ariane Pinche, Matthias Gille Levenson, Margot Ferrand. COSME^2 - Complexities: 30 Years Of Research Of Medievalists DH Concerning A Thousand Years Of Medieval Sources. Digital Humanities 2019 : Complexities, Jul 2019, Utrecht, Netherlands. hal-03156820

HAL Id: hal-03156820
https://hal.science/hal-03156820

Submitted on 30 Oct 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.

COSME²

Complexities: 30 Years Of Research Of Medievalists DH Concerning A Thousand Years Of Medieval Sources

A first statement of fact: the lack of compability between corpuses because of their complexity.

In the last 30 years, a lot of digital editions and corpora have been produced; yet they are often uncompatible with one another, and they are quite difficult to parse.
Nowadays, projects to edit and publish sources are very numerous, but very often scattered; in a word, there is a lack of connection between them.
Besides, it is difficult to deal with the complexity of sources, and to connect corpora of different nature, between archive documents and narrative sources, for example.

A solution: interoperability

Cosme² (**Co**nsortium **S**ources **Mé**diévales) was created to **gather medievalists around their sources**. Its main aim is to **connect corpora**.

We think that **data interoperability** can be a way to achieve this:
- by the digitalisation of non digital editions;
- by the structuration of data of the digital native / non native corpora (TEI, METS, EAD);

TEI – TEXT ENCODING INITIATIVE

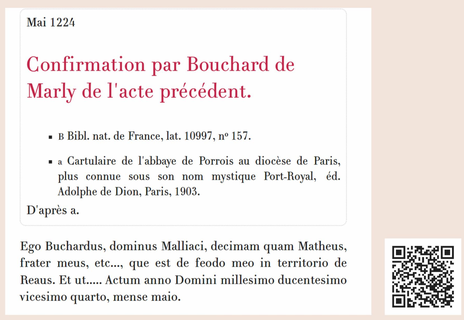
<ead> Encoded Archival Description

METS Metadata Encoding & Transmission Standard Official Web Site

- by the processing of heterogeneous sources: *de facto* interoperability (e.g., work on the collection of named entities from multiple corpora).

The corpora treated represent a **large number of disciplines** (history, art history, archaeology, literature, linguistics) **and supports** (text, objects, image).

Cuartillo representing Henri IV of Castille's motto:
A grenade with its flowers and fruits
1450-1480
DEVISE database, item n°901.
Université de Poitiers



Mai 1224

Confirmation par Bouchard de Marly de l'acte précédent.

- B Bibl. nat. de France, lat. 10997, nº 157.
- a Cartulaire de l'abbaye de Porrois au diocèse de Paris, plus connue sous son nom mystique Port-Royal, éd. Adolphe de Dion, Paris, 1903.

D'après a.

Ego Buchardus, dominus Malliaci, decimam quam Matheus, frater meus, etc..., que est de feodo meo in territorio de Reaus. Et ut..... Actum anno Domini millesimo ducentesimo vicesimo quarto, mense maio.

Cartularium from Port-Royal-des-Champs, in "Cartulaires d'Île-de-France", Telma, École des Chartes.

COSME²:

- is active since 2013, with Cosme¹
- collaborates with more than 20 research teams
- has 15 researchers directly involved
- maintains a website: cosme.hypotheses.org

Our ways to get into interoperability:

1 · Animation of working groups:

- Lemmas

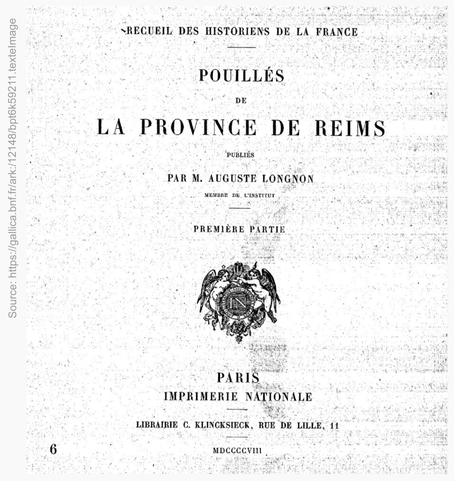
Cosme² is particularly interested in lemmatisation. Various projects are oriented towards the creation of lemmatised corpora of various nature (inscriptions, litterary texts, historical sources). There is no consensus on the tools to use nor on the tag sets: the WG 'lemma' gathers 6 research teams across three countries (Belguim, France, Germany) to address this problem.

Form	Lemma	POS	Morph
Signeur	Seignor	NOMcom	NOMB.=s GENRE=m CAS=r

Pyrrha, interface of lemmatisation post-correction, produced by the École nationale des Chartes, ***https://dh.chartes.psl.eu/pyrrha/***

- Shelfmarks

Work on shelfmarks with the scriptorium.be manuscripts database (more than 20.000 entries). Reflexions on the implementation of an **ISO standard** for manuscripts identifier, the ISMI (International Standard Manuscript Identifer)



Source: https://gallica.bnf.fr/ark:/12148/bpt6k59211.texteImage

- PlaceNames / PersNames

Creation of a national (french) referential of placenames: XML encoding of the indexes of 2 volumes of the Pouillés de la France; after this, a collection of the toponyms is being made, to create a JSON datamodel, to strengthen interoperability.
This group also works on personal names, focusing its efforts on prosopography.

2 · Organization of workshops

Cosme² is willing to direct part of its resources towards pedagogy. We aim to train academics to use new tools, methods and standards that are in use when working with medieval sources. A training session on XSLT, in Lyon, and a workshop on lemmatisation that took place in the IRHT were organised this year.

3 · Funding and co-funding of projects

Telma, PALM, SCRIPTA, TRANSSCRIPT, etc.

Telma

Telma gathers a set of digital editions of medieval historical sources, and offers assistance to researchers aiming to create digital editions and databases. The result is a corpus of databases that can be "cross-queried". *http://www.cn-telma.fr/*

4 · Publication (work in progress)

Cosme² plans to publish a white paper on creation and management of databases (planned for 2020), centered on the production workflow, and based on the guidelines of Huma-Num, an important Research Infrastructure (Très Grande Infrastructure de Recherche).

Download this poster:

Huma-Num

cnrs

Contact us: