



Workshop Serica. Pisa, 6 giugno 2022

Implementazioni della piattaforma Eurasian Latin Archive



Fase 1

2018-2020

DFCLAM	P.I. Francesco Stella Costanza Marzucchi Elisabetta Bartoli, Irene Volpi
DIISM	Monica Bianchini Nicola Giannelli
QUESTIT	Leonardo Rigutini
UniSI- QuestIT	Emmanuela Carbé

Fase 2

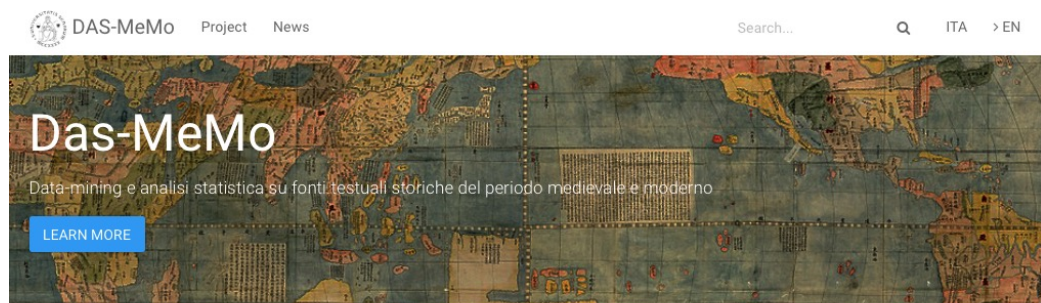
2021-2022

DFCLAM	P.I. Francesco Stella Emmanuela Carbé Silvia Bettella (2020)
QUESTIT	Alessandro Gronchi Nicola Giannelli Chiara Pugliese



Fase 1

marzo 2018-febbraio 2020



Eurasian Latin Archive

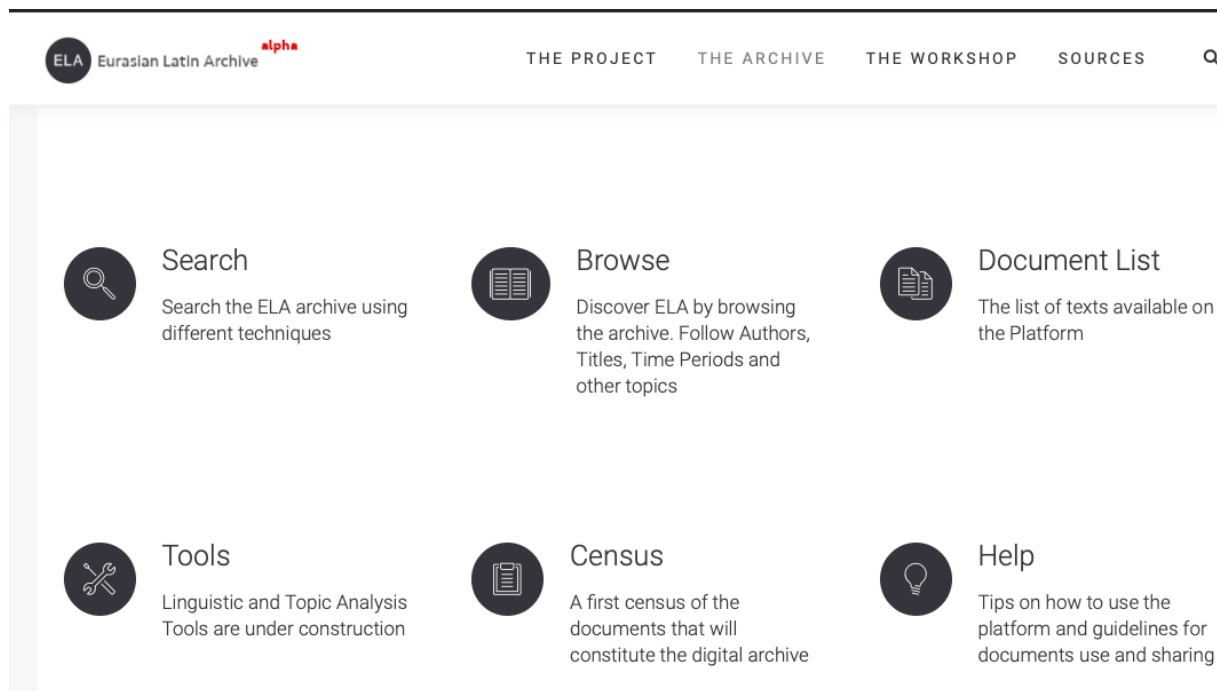
The "Eurasian Latin Archive" project was born for the creation of a digital library consisting of Latin texts and documents from medieval and early modern times concerning East Asia, including a specific "Latin Silk Road" section dedicated to Latin or Latin-Chinese texts.

The Eurasian Latin Archive will be provided with digital mining tools for textual and thematic analysis. Aim of the collection is the comparative language analysis, both internal with other latin texts from different eras and areas, and with non-latin texts with homogeneous subjects.

Materials will be available via a platform that includes tools for language and semantic analysis, which is in the design phase to date within the Das-MeMo project, started March 2018. The



dasmemo.unisi.it



ela.unisi.it

- Censimento di testi candidati per entrare nell'Eurasian Latin Archive
- Gestito con OpenRefine
- Contiene metadati descrittivi , strutturali e amministrativi delle risorse (poi in <teiHeader>): informazioni su autore/i (VIAF/Wikidata), titolo, eventuale riferimento all'edizione dell'opera e pp. (Worldcat), link a Bibliotheca Sinica 2.0 e CCT-Christian Texts Database
- Il censimento funge anche da gestionale per il lavoro di digitalizzazione, trascrizione e codifica in XML TEI dei testi: i collaboratori che prendono in carico un documento aggiungono progressivamente all'item le indicazioni del processo (nome editor, ORCID, data di inizio del lavoro, punteggio sulla qualità e attendibilità della trascrizione o OCR (da 1 a 3), qualità della codifica TEI (da 1 a 3), eventuale software utilizzato per gli OCR, data di pubblicazione in ELA, indicazione sui diritti di utilizzo del documento, eventuali modifiche dopo la prima versione (data, tipo di modifica, identificativo ORCID del responsabile della modifica)

46	30/09/18	Hieronymus Mayoric	Tertius catalogus Provinciae Japonicae 1654	ABBY FineReader 12. Catalogi Provinciae . Monumenta His			https://
48	30/09/18		Elenchus patrum ad congregationem Pri 1654	ABBY FineReader 12. Catalogi Provinciae . Monumenta His			https://
49	30/09/18	Francisco de Xavier	Francisco de Xavier a Ignacio de Loyola 1548	ABBY FineReader 12.1.11	Monumenta His 1699-1783	https://	https://www.worldcat.org/
50	30/09/18	Francisco de Xavier	Francisco de Xavier a Ignacio de Loyola 1548	ABBY FineReader 12.1.11	Monumenta His 1699-1783	https://	https://www.worldcat.org/
51	30/09/18	Francisco de Xavier	Francisco de Xavier a Ignacio de Loyola 1549	ABBY FineReader 12.1.11	Monumenta His 1699-1783	https://	https://www.worldcat.org/
52	30/09/18	Francisco de Xavier	Francisco de Xavier a Simon Rodrigues 1549	ABBY FineReader 12.1.11	Monumenta His 1699-1783	https://	https://www.worldcat.org/
53	30/09/18	Ignacio de Loyola	Ignacio de Loyola a Francisco de Xavier 1549	ABBY FineReader 12.1.11	Monumenta His 1491-1556	https://	https://www.worldcat.org/
54	30/09/18	Ignacio de Loyola	Ignacio de Loyola a Francisco de Xavier 1549	ABBY FineReader 12.1.11	Monumenta His 1491-1556	https://	https://www.worldcat.org/
55	30/09/18	Melchor Nuñez Barreto	Melchor Nuñez Barreto a Ignacio de Lo 1555	ABBY FineReader 12.1.11	Monumenta His 1520-1571	https://	https://www.worldcat.org/
56	30/09/18	Melchor Nuñez Barreto	Melchor Nunes Barreto a Diego Lainez 1558	ABBY FineReader 12.1.11	Monumenta His 1520-1571	https://	https://www.worldcat.org/
57	30/09/18	Juan de Polanco	Anotaciones sobre las misiones 1558	ABBY FineReader 12.1.11	Monumenta His 1516?-1576?	https://	https://www.worldcat.org/
58	30/09/18	Diego Lainez	Normas Particulares para la India // De 1558	ABBY FineReader 12.1.11	Monumenta His 1512-1565	https://	https://www.worldcat.org/
59	30/09/18	Nicolaus Lancillottus	P. Nic. Lancillottus S. I. P. Ignatio de L. 1546	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
60	30/09/18	Ignacio de Loyola	P. Ignatius De Loyola P. Paulo S.I. Goa 1547	ABBY FineReader 12.1.11	Monumenta mis 1491-1556	https://	https://www.worldcat.org/
61	30/09/18	Ioannes de Polanco	P. Ioannes de Polanco S.I. ex Comm. P. 1547	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
62	30/09/18	Ioannes de Polanco	P. Ioannes de Polanco S.I. ex Comm. P. 1547	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
63	30/09/18	Ioannes de Polanco	P. Ioannes de Polanco S.I. ex Comm. P. 1548	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
64	30/09/18	Antonius Criminalis	[P. Antonius Criminalis S.I.] P. Ignatio 1549	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
65	30/09/18	Ignacio de Loyola	Ignatius De Loyola P. Francisco Xavei 1550	ABBY FineReader 12.1.11	Monumenta mis 1491-1556	https://	https://www.worldcat.org/
66	30/09/18	Nicolaus Lancillottus	P. Nicolaus Lancillottus S.I. P. Ignatio c 1550	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
67	30/09/18	Ludovicus Gonçalves da Camara	P. Ludovicus Gonçalves da Camara S.I. 1550	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
68	30/09/18	Ignacio de Loyola	Ignatius De Loyola Francisco Xaverio c 1550	ABBY FineReader 12.1.11	Monumenta mis 1491-1556	https://	https://www.worldcat.org/
69	30/09/18	Nicolas Trigault (金尼閣)	P. N. Trigaulti Relatio de P. Gasparis B. 1550	ABBY FineReader 12.1.11	Monumenta mis 1577-1628	https://	https://www.worldcat.org/
70	30/09/18	Nicolaus Lancillottus	P. Nicolaus Lancillottus S.I. P. Ignatio c 1551	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
71	30/09/18	Ioannes de Polanco	P. Ioannis De Polanco S. I. Excerpta e li 1556	ABBY FineReader 12.1.11	Monumenta mis	https://	https://www.worldcat.org/
72	07/01/19	Alexander Valignano	Catechismus christianae fidei editus a Pa 1586	Transcribed	1539-1606	https://	https://
73	07/01/19	Prospero Intorcetta	et al. Sapientia sinica, exponents P. Ignatio a 1662	Fondazi Transcribed (2 pages)	1625-1696	https://	https://www.
74		Prospero Intorcetta	Sinarum scientia politico-moralis 1667	Fondazione Intorcetta	1625-1696	https://	https://www.
75		Prospero Intorcetta	Confucius Sinarum philosophus, sive sc 1687	Fondazi http://www.fondazioneintorcetta.info/pdf/Confucius	1625-1696	https://	https://www.
76		Prospero Intorcetta	Compendiosa narratio. De statu mission 1671	Fondazi http://www.fondazioneintorcetta.info/pdf/Compendio	1625-1696	https://	https://www.
77		Prospero Intorcetta	De cultu sinensi, datum anno 1668 1668	Google Books	1625-1696	https://	https://www.

Fase 1



- I processi si differenziano in base alla situazione di partenza, una risorsa ad esempio può essere già stata digitalizzata da altre istituzioni e pubblicata online. Il lavoro può comprendere digitalizzazione, OCR e controllo, trascrizioni, codifica TEI, adattamento della TEI di un testo già codificato da altre istituzioni
- Testo codificato in XML TEI P5, modello simile a ALIM
- Semplificazione di alcuni livelli di codifica strutturale e utilizzo di elementi di codifica semantica
- Lo scioglimento delle abbreviazioni (<abbr><expan>) migliora le prestazioni del lemmatizzatore
- <persName> con VIAF o Wikidata
- <placeName> con Pleiades o GeoNames
- <foreign>: marcatura degli inserti in lingua non latina (ISO 639), compreso il Pinyin; sono marcate come lang non ISO, sebbene non siano propriamente lingue, alcune traslitterazioni o rese di pronuncia in latino del cinese e giapponese
- realizzati alcuni script in Python per l'inserimento automatico dei tag in occorrenze già individuate: processano un file .csv arricchito con OpenRefine (VIAF, Wikidata, Pleiades, GeoNames)

Fase 1



ELA TOOL

LA Tool è un framework basato su CLTK e per il trattamento dei testi ELA in latino e multilingua.

Svolge le seguenti operazioni, restituendo i risultati in formato JSON:

Parsing <teiHeader>

Estrazione metadati

Parsing <text>

Normalizzazione e cleanup

Tokenizzazione

[POS]

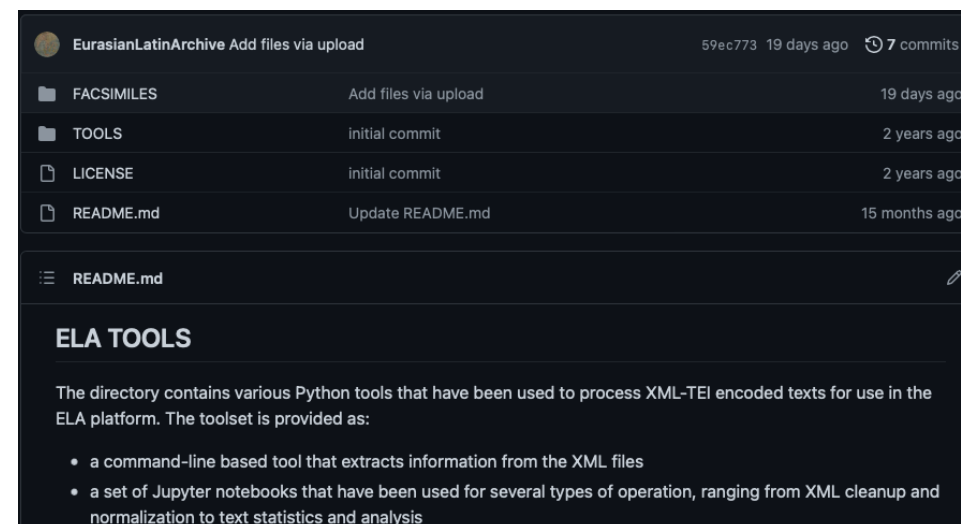
Lemmatizzazione latino (backoff method)

Statistiche

Collocations

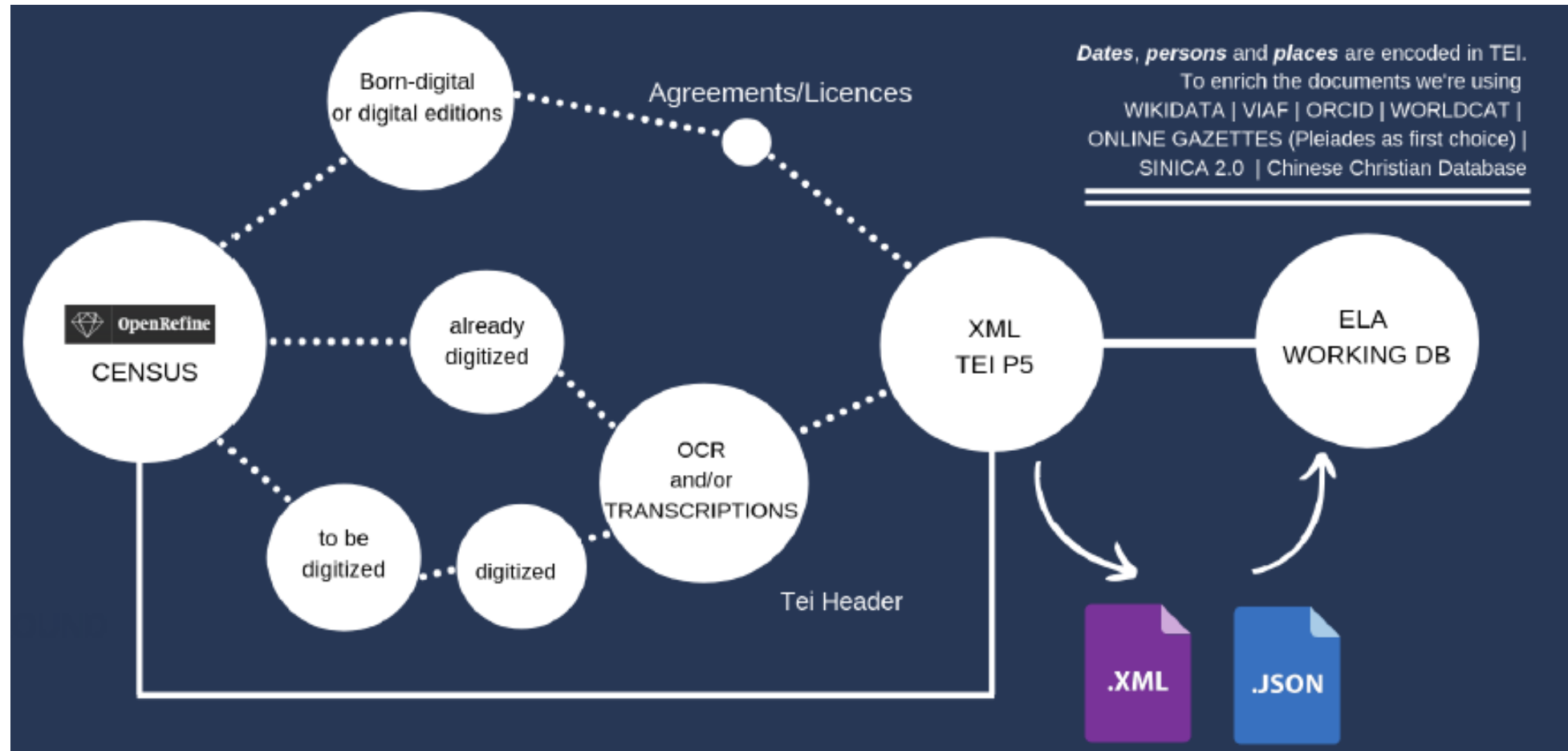
Word N-Grams

Estrazione coordinate geografiche da Pleiades e GeogNames



<https://github.com/EurasianLatinArchive/ELA>

Fase 1



Fase 1



BACKEND

- Backend Java EE
 - Database MySql
 - ElasticSearch
-
- Estrazione e mappatura della struttura del documento TEI e della lemmatizzazione di ELA Tool
 - Permette ricerche full text (sintassi Lucene), per lemma, per tag, con filtri
 - Gestione utenti e ruoli (admin, editor, reviewer, user non autenticato)
 - Gestione documento (id, upload, modifica, cancellazione, versioni)
 - Esportazione documento (PDF/XML/TXT)
 - Gestione versioni di ELA Tool: refresh dei testi quando è disponibile una nuova versione del tool

Fase 1



INTERFACCE

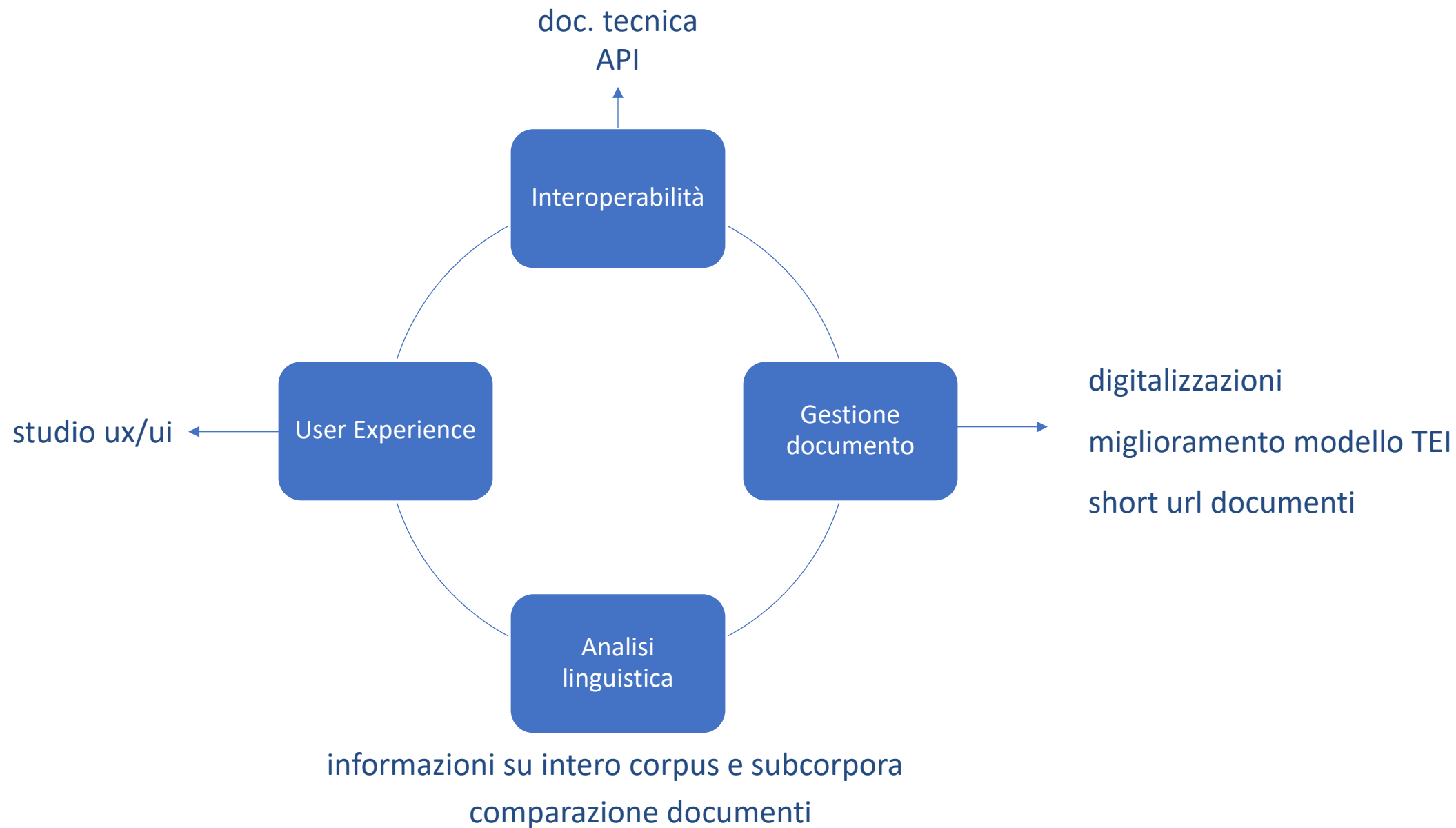
(1) Interfaccia Generic User

- CMS WordPress
 - Presentazione progetto
 - Guida per l'utente
 - Censimento
 - Notizie
- Motore di ricerca (Javascript):
 - Search (ricerche full text con sintassi Lucene, per lemma, con filtri)
 - 'Browse by' (tag)
 - Lista semplice dei testi
- Risultati di ELA Tool (Javascript)
- Mappe (Leaflet)
- Visualizzatore Tei Boilerplate

(2) Interfaccia Content Creator

- Autenticazione
- Ricerca full text
- Caricamento/eliminazione item
- Refresh ELA Tool

Fase 2



Fase 2

Document List

Search:

Author	↑↓	Title	↕	Date	↕	Share	View	Info
Andreas de Perusio O.F.M.		Epistola		1326				
Augustinus Sancti Paschali O.F.M.		Responsum ad intimationem D. Maigrot 30 nov. 1689		1689				
Augustinus Sancti Paschali O.F.M.		Responsum ad Procuratorem Generalem, 11 Dec. 1689		1689				
Augustinus Sancti Paschali O.F.M.		Epistola ad Illm. Bernardinum Della Chiesa, 15 Dec. 1684		1684				
Augustinus Sancti Paschali O.F.M.		Responsa ad D. Carolum Maigrot, 15 et 22 Mart. 1689		1689				

Epistola

Andreas de Perusio O.F.M. 12..-1332

in: Sinica Franciscana, I: Itinera et relationes fratrum Minorum saeculi XIII et XIV, a cura di A. Van Den Wyngaert, Firenze, Quaracchi, 1929, 373-377

page: 373

Fr. [Andreas de Perusio](#) de Ordine Minorum fratrum divina permissione vocatus Episcopus, reverendo patri fratri Guardiano [Perusini Conventus](#) salutem et pacem in Domino sempiternam. Et infra sequitur.

1. Nam propter immensam terrarum mariumque distantiam, inter me et vos interiectam, vix sperare possum, quod littere ad vos per me transmissae ad manus vestras valeant pervenire. Et infra subsequitur. Noveritis me itaque cum bone memorie fratre [Peregrino Coepiscopo](#), et mee peregrinationis individuo comite, post multos labores et langores, inedia, variaque incomoda, atque pericula in terra pariter et in mari, ubi fuimus rebus omnibus et etiam tunicis et habitibus spo

page: 374

liati, demum, Deo iuvante, ad [Cambaliensem civitatem](#), que sedes est imperii [magni Chanis](#), [anno dominice incarnationis millesimo CCC° XVIII](#), ut credo, pervenisse; ubi secundum mandatum a [Sede Apostolica](#) nobis datum, Archiepiscopo consecrato, moram ibi per quinquennium ferme contraximus. Infra quod temporis spatium procuravimus alafa ab Imperatore magnifico pro victu et vestitu octo personarum. Est autem alapha impense, quas Imperator tribuit nunciis magnatorum, oratoribus, bellatoribus et diversarum artium artificibus et ioculatoribus, pauperibus et diversis diversarum conditionum personis; que impense plurium latinorum Regum introitus expensasque transcendunt.

2. De divitiis, magnificentia et gloria huius magni Imperatoris, de vastitate Imperii, multitudine populorum, numerositate civitatum, et magnitudine earumdem, et de ordinatione imperii, in quo nemo adversus alium ausus est levare gladium, transeo, quia longum foret

Fase 2

Search

Text

?

Lorem Ipsum Dolor Sit Amet

Lemmas

?

Filters

?

+

Search

Full corpus

Results

Show 5 entries

Author	Title	Date	Compare		Share	View	Info
Andreas de Perusio O.F.M.	Epistola	1326					
Augustinus Sancti Paschali O.F.M.	Responsum ad intimationem D. Maigrot 30 nov. 1689	1689					
Augustinus Sancti Paschali O.F.M.	Responsum ad Procuratorem Generalem, 11 Dec. 1689	1689					
Augustinus	Epistola ad						

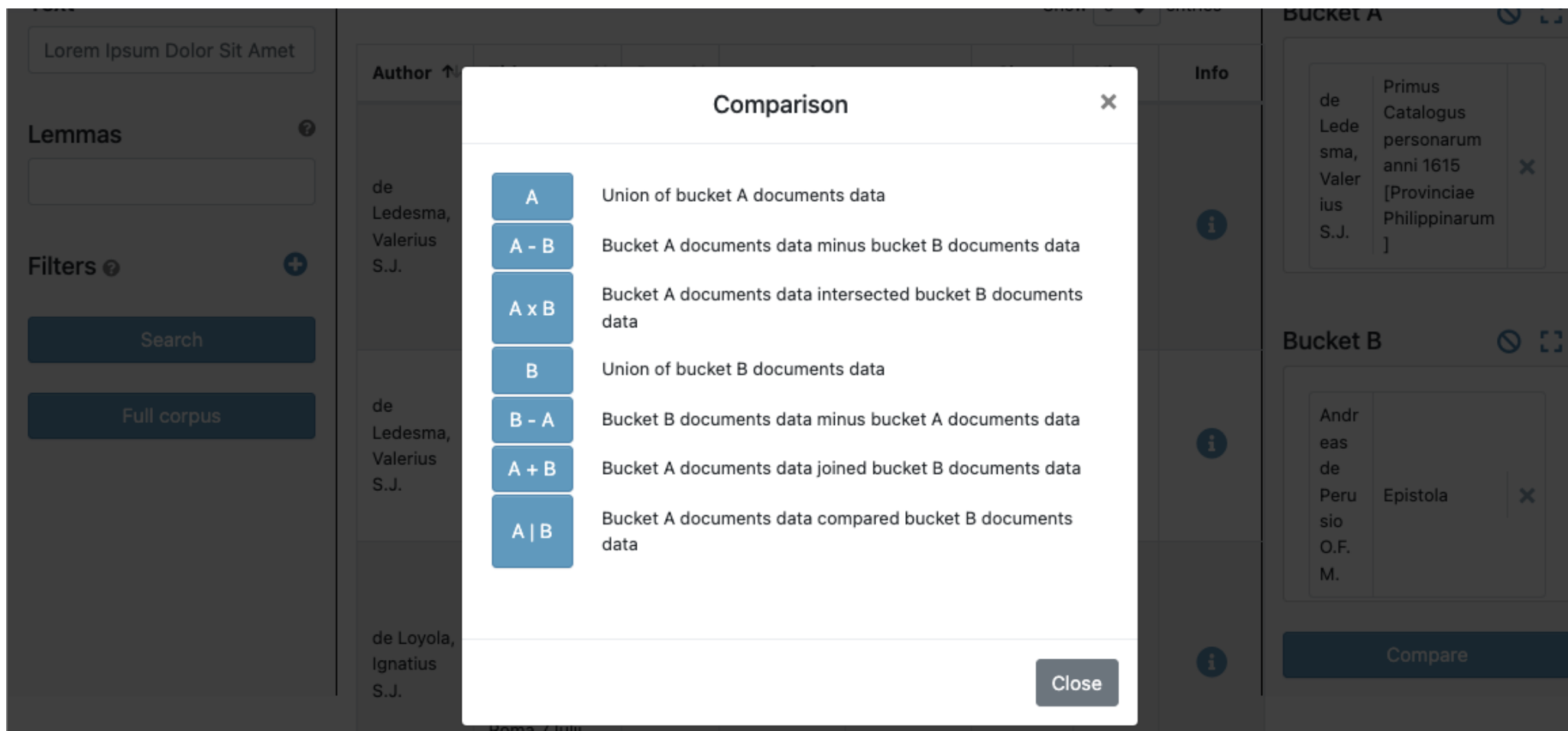
Compare

Bucket A

Bucket B

Compare

Fase 2



Fase 2

Comparison info



Stats

Words

Latin Words

Latin Stop Words

Lemmas

Persons

Places

Show

10

 entries

Search:

Number of words	198
Number of Latin words without stop words	123
Numbers of Latin lemmas	249

Showing 1 to 3 of 3 entries

Previous

1

Next

Download all data

Download current table

Fase 2

A

A - B

A x B

B

B - A

A + B

A | B

Union A

Stats

Words

Latin Words

Latin Stop Words

Lemmas

Persons

Places

Show 10 entries

Search:

Number of words	6648
Number of Latin words without stop words	5205
Numbers of Latin lemmas	6648

Showing 1 to 3 of 3 entries

Previous1Next

Download all data

Download current table

Union B

Stats

Words

Latin Words

Latin Stop Words

Lemmas

Persons

Places

Show 10 entries

Search:

Number of words	3749
Number of Latin words without stop words	2996
Numbers of Latin lemmas	3749

Showing 1 to 3 of 3 entries

Previous1Next

Download all data

Download current table