

The Library of Congress
LC's Digital Future and You!
a series of briefings sponsored by Library Services on digital initiatives

Tuesday, February 6, 2018

The SHARE-VDE Project: Fulfilling the Potential of BIBFRAME

Michele Casalini
Casalini Libri

Tiziana Possemato
Casalini Libri - @Cult



Index

- Introduction and overall project goals
- Theoretical context
- SHARE-VDE process overview
- Entity identification, Reconciliation and Data enrichment
- Reconciliation & Enrichment - Automatic procedures
- Reconciliation & Enrichment - Manual procedures
- Conversion in RDF/BIBFRAME
- Trust and Provenance
- SHARE-VDE phase 2 deliverables overview
- SHARE-VDE possible steps for the future



Introduction and overall project goals



Current activity and infrastructure

Casalini Libri produces, for publications from Romance language countries, more than 40,000 original bibliographic records in RDA as a member of the Program for Cooperative Cataloguing (PCC) with authority entries;

Bibliographic records are created using the @Cult OLISuite WeCat cataloguing modules;

@Cult, in addition to the LMS and Discovery tools field, is specialized in the development of software components and platforms to convert, enrich, reconcile and publish the data of cultural institutions under the linked data paradigm.



The three major areas of activity towards the BIBFRAME/Linked Data environment

The enrichment of MARC records with URIs to simplify conversion into BIBFRAME;

The use of a framework to automate the conversion from MARC to RDF, using BIBFRAME vocabulary;

The creation of a BIBFRAME layered platform prototype starting from bibliographic and authority records, to test and demonstrate the advantages of the BIBFRAME data model.



SHARE-VDE overall goals

The main goals of Phase 1 and 2 Research & Development activities are:

- Reconciliation and clusterization of variant forms of the same entity;
- Enrichment of MARC records with URIs, with the development of detection procedures for entity identification, including relator terms;
- Conversion, supply and management of authority and bibliographical data in BIBFRAME, taking into account the complexity of the long and heterogeneous transition time both for libraries and data producers;
- Publication of a BIBFRAME three-layered platform prototype.



SHARE Virtual Discovery Environment project

Each participant decides whether or not to take part in the subsequent phases.

Phase 1: analysis, enrichment, reconciliation, conversion in RDF, publication of two sets of bibliographic data for each participating library were foreseen (1985 and 2015 imprint titles). This phase also included the release of MARC records enriched with URIs and BIBFRAME 1.0 datasets for each participating library.

A total of 2,249,387 bibliographical records and 3,601,327 authority records were converted into BIBFRAME 1.0 and published on the SHARE-VDE portal.

Phase 1: from October 2016 to January 2017.



SHARE Virtual Discovery Environment project

Phase 2: data enrichment and conversion refinements and customization, enhanced data supply workflow experimentation, second release of the portal.

The library catalogue of each participating institution is to be converted into BIBFRAME 2.0 and returned to each library (over 100 million records and consequent datasets are expected to be processed).

A relationship database that registers the relationships between entities (person, work, instances, subjects, publisher, ...) is established in order to assure a more precise identification rate of each entity and reach a higher quality of results without human intervention.

Refinement of data, e.g. for co-authors and editors, where they are identified in a variety of ways within library records (Relator terms topic).



SHARE Virtual Discovery Environment project

... cont. Phase 2:

Export of data in MARC or RFD format filtering the library preferred URIs.

Inclusion of additional URI sources, e.g. specific source for corporate bodies, subjects (LCSH, FAST ...), RDA vocabularies.

Analysis for the creation of relationships among subject terms and strings in different languages.

Provenance declaration, update management and built-in instances will be addressed.

Phase 2: from March to December 2017.

Phase 3: modular implementation in production of the various components, according to the specific workflow of the each library.



Participating libraries (1)

Phase 1	Phase 2	(in Country/State order):
X	X	Stanford University
X	X	University California Berkeley
X	X	Yale University
X	X	Library of Congress
X	X	University of Chicago
X	X	University of Michigan Ann Arbor
X	X	Harvard University
X		Massachusetts Institute of Technology
	X	Duke University
X		Cornell University
X		Columbia University
X	X	University of Pennsylvania



Participating libraries (2)

Phase 1	Phase 2	(in Country/State order):
	X	Pennsylvania State University
X	X	Texas A&M University
	X	University of Alberta
X		University of Toronto

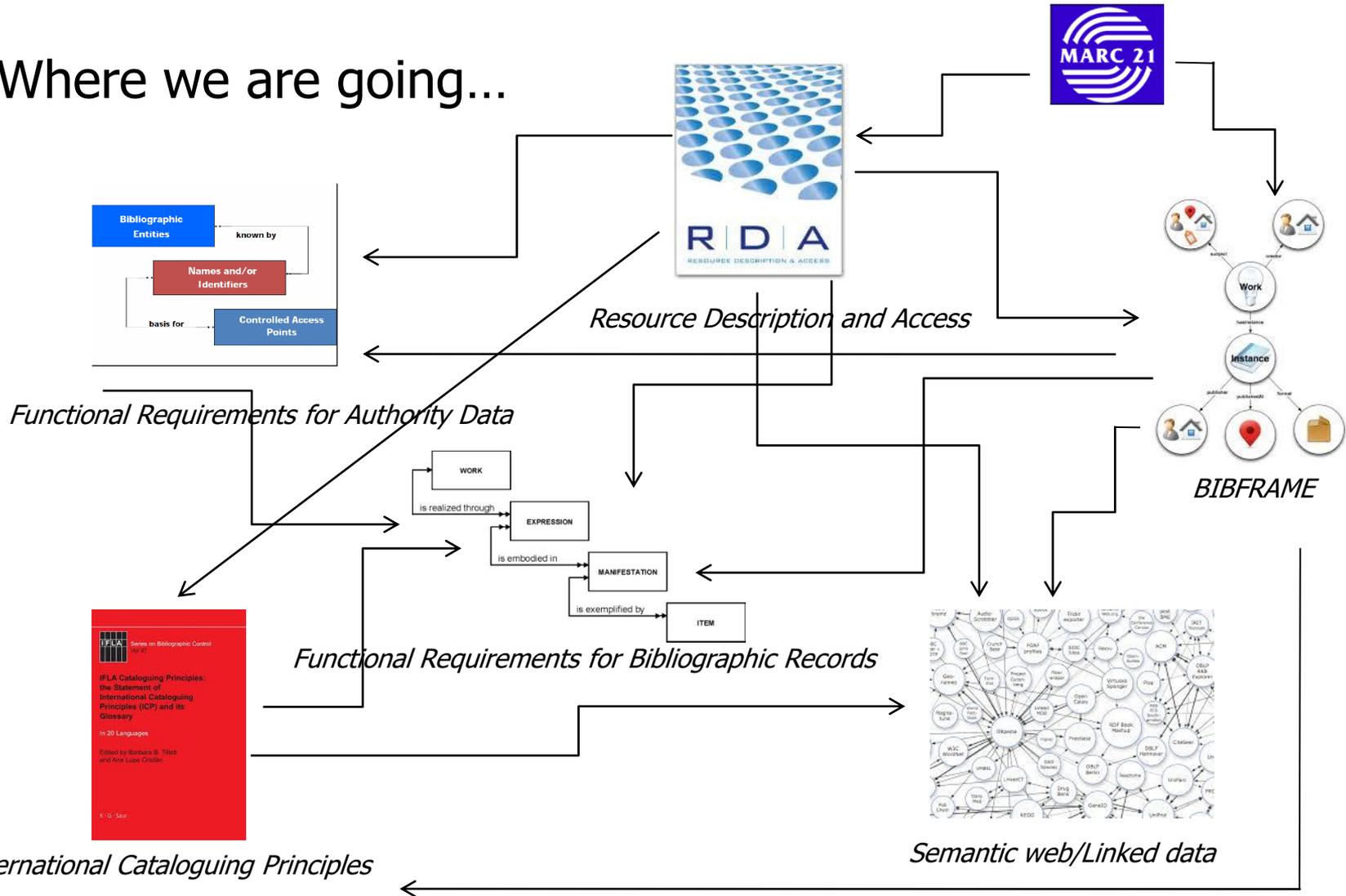


Theoretical context



The theoretical context of the project

Where we are going...



The theoretical context of the project

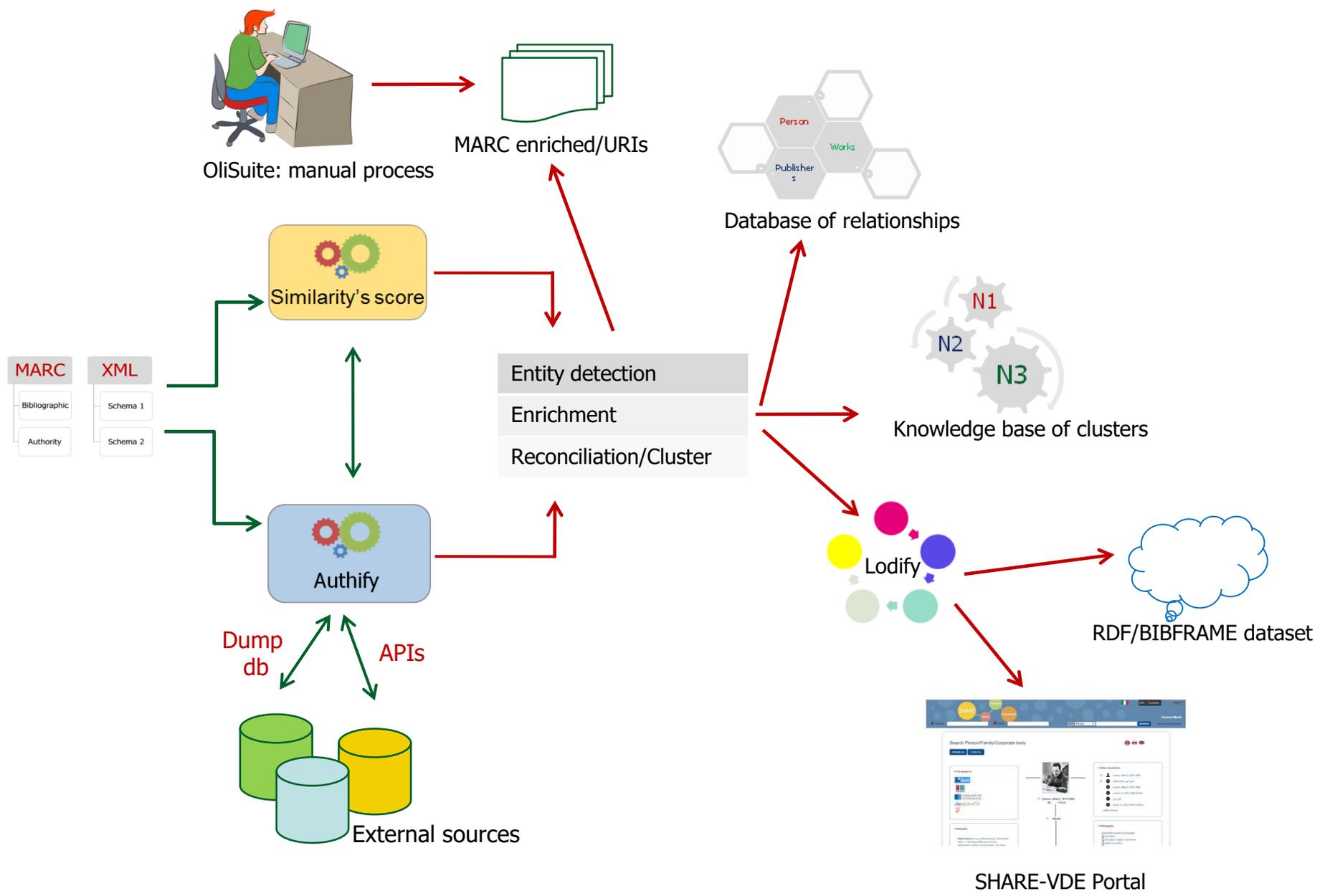
New standards, models and technologies as ways to approach entity **identification** and the **relationships** between entities, recognized as the key element in the construction of new **entity detection** and **entity identification** processes:

- **RDA – *Resource Description and Access***, the new international guidelines to manage resources
- **Linked Open Data** philosophy and technology
- **BIBFRAME**: one of more interesting models to convert and publish data. This model is considered ‘the core’ ontology, completed with the ontologies for specific domains, that libraries will suggest

SHARE-VDE process overview



The SHARE-VDE processes

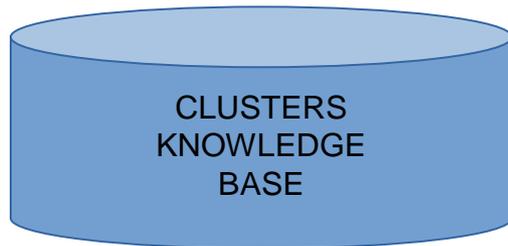
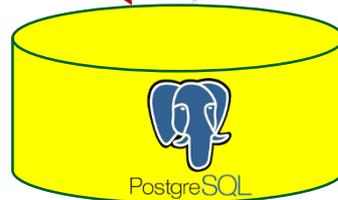
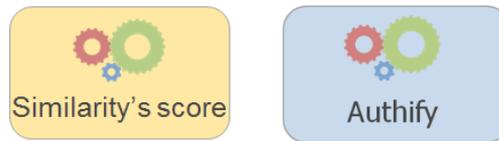


Focus on processes 1/2

Authority records



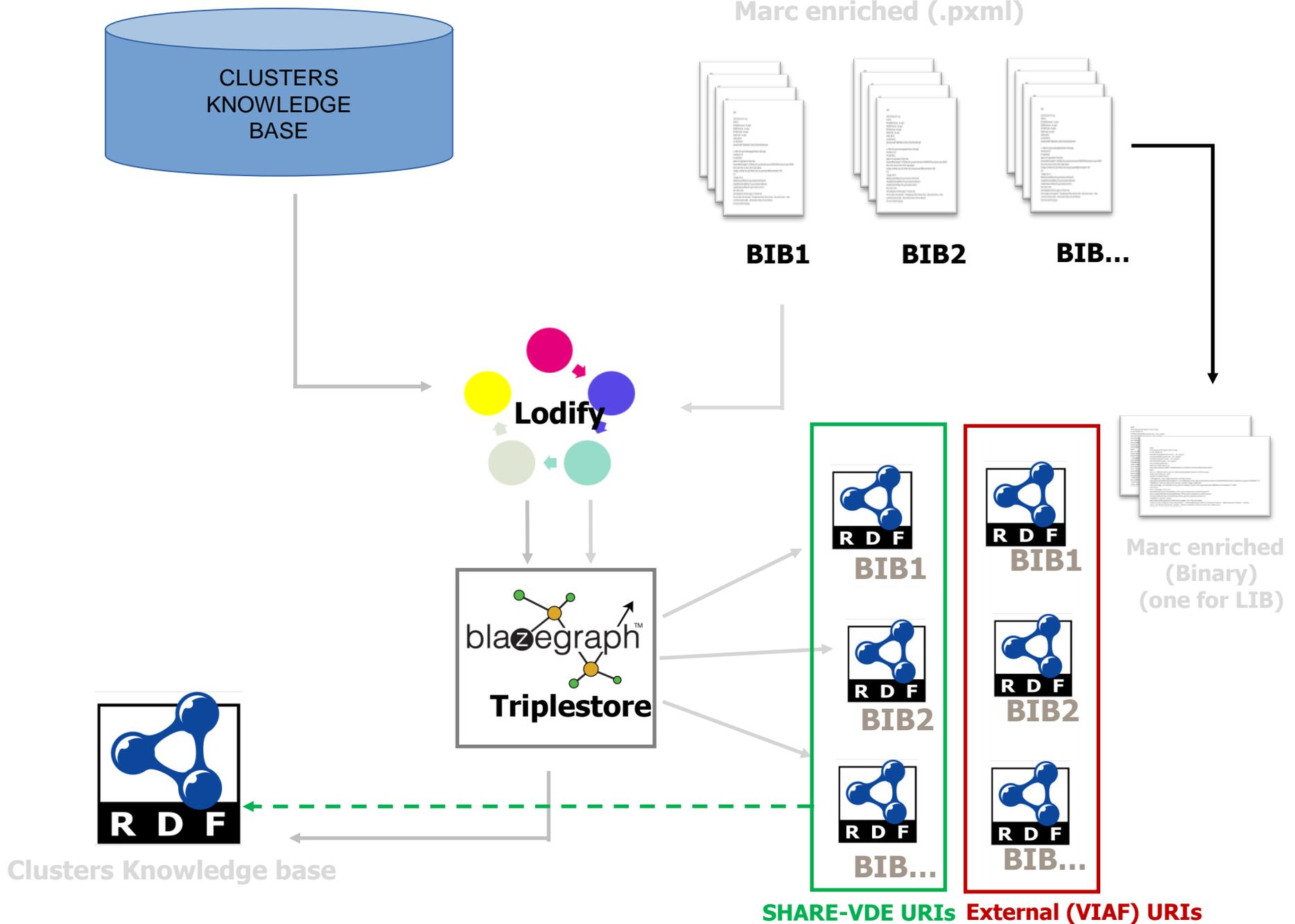
Bibliographic records



Marc enriched (.pxml)



Focus on processes 2/2



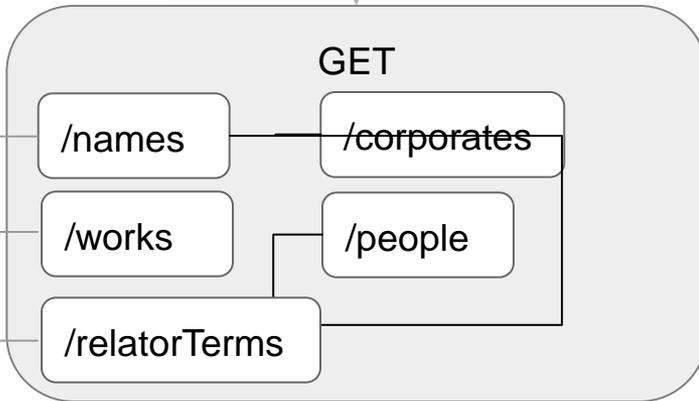
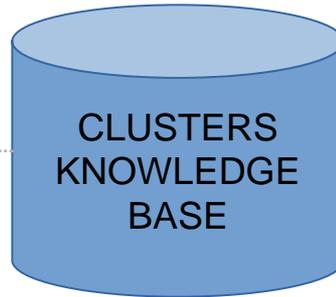


Injection services (massive)

Injection services (massive)

Cluster search services

Injection services (single cluster)



VIAF

110,000,000 Bibliographic Records

19,200,000 Names

20 Million Page Requests in 2011



LCSH

stands for

Library of Congress Subject Heading



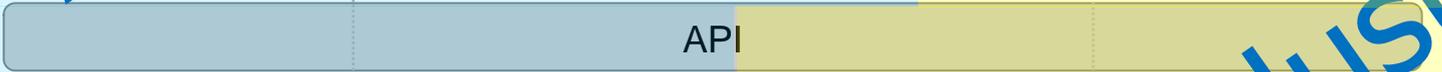
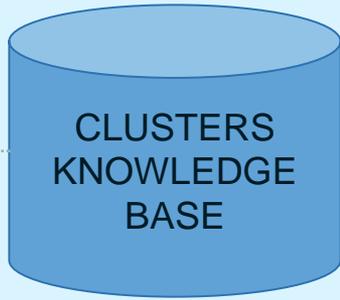
Injection services (massive)

Injection services (massive)

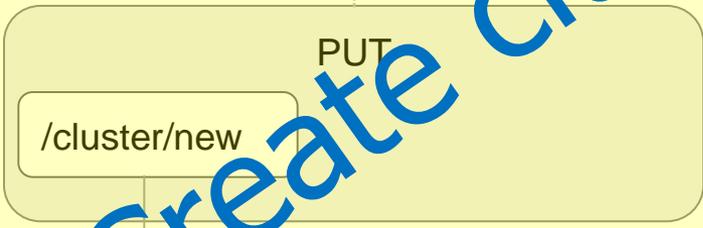
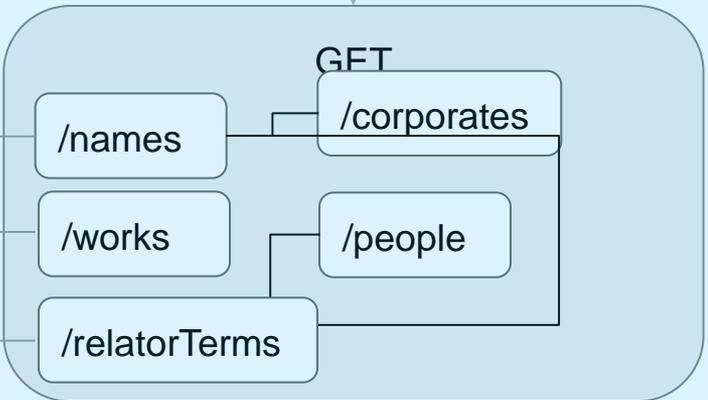
Authify

Cluster search services

Injection services (single cluster)

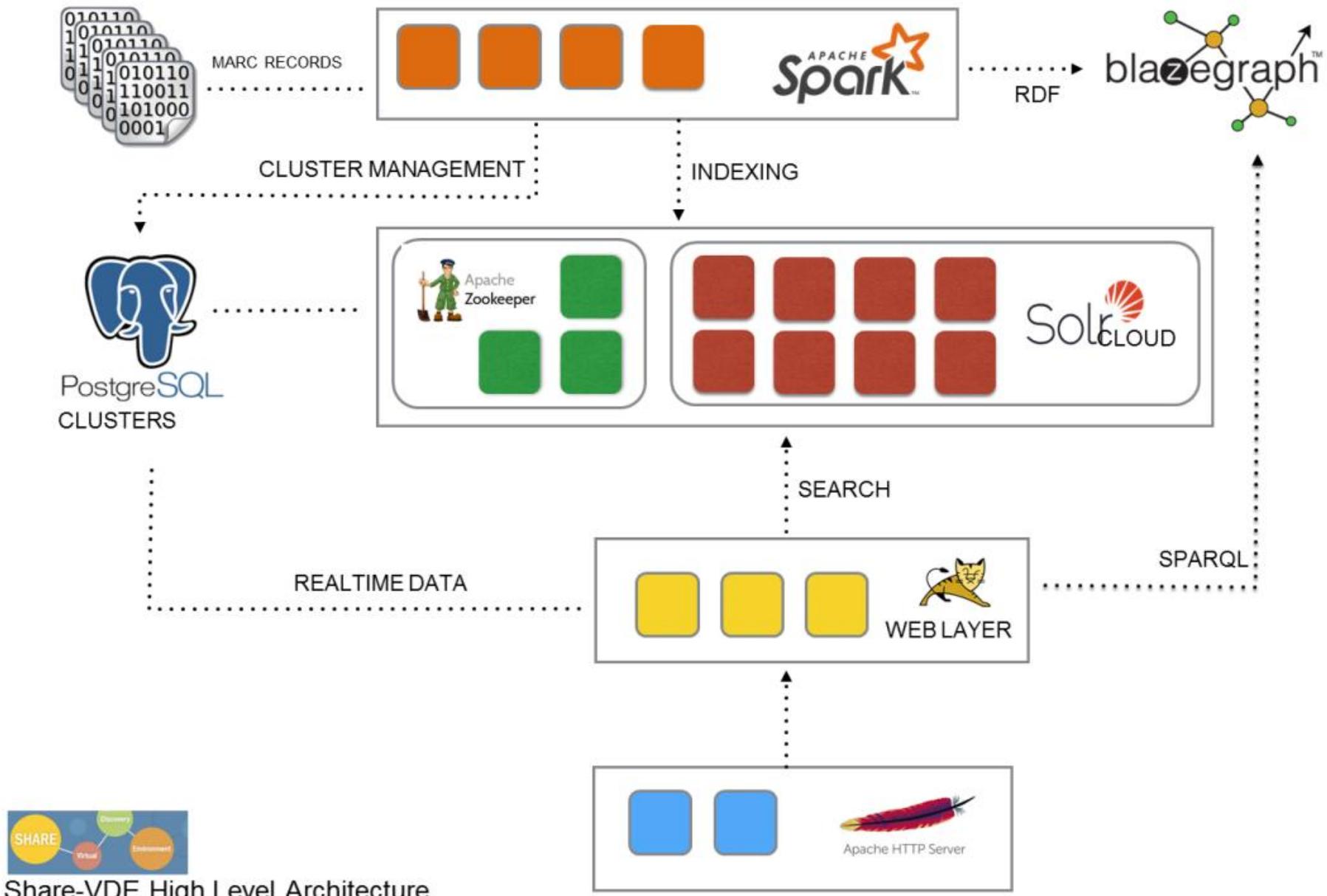


API



Create cluster





Share-VDE High Level Architecture

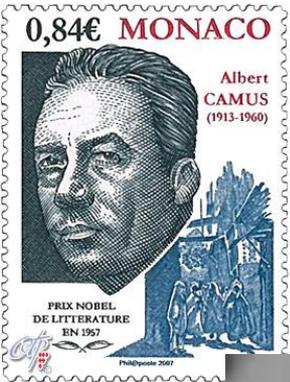
Entity identification, Reconciliation and Data enrichment



Who's Who?

The question at hand:
how to identify an entity?





I don't care, and you find that attractive. But I also don't care about the



Happy Valentine's Day ♡
Love, _____ and
Albert Camus
benkling.tumblr



Albert Camus



http://share-vde.org/sharevde/searchNames?n_cluster_id=133656

The importance of identification in the cataloguing tradition (and not only!)



Entity identification: it has traditionally been considered a highly important aspect of cataloguing.

But, the use of attributes to identify a person has not been widely used

* *Both pictures are taken at the City Lights Bookstore, in San Francisco*

Data reconciliation and enrichment

With the online presence of different catalogues and authority files available in various formats and, where possible, in open mode, the concepts of authority control and of union catalogue have also evolved into the **grouping of an entity's identifying attributes** from different sources.

The process is best known as ***reconciliation*** and consists in **creating a cluster of data that all refer to the same entity**.



The new revolution: from record to entity

Full Record	MARC Tags
	Where to Request
000	02384cam a2200445 a 4500
001	15789102
005	20100204124659.0
008	090624s2009 enka b 000 0 eng
010	__ a 2009025024
015	__ a GBA959347 2 bnb
016	7_ a 015291646 2 UK
020	__ a 9780521519748 (hbk.)
020	__ a 0521519748 (hbk.)
020	__ a 9780521732505 (pbk.)
020	__ a 0521732506 (pbk.)
035	__ a (OCoLC)ocn416716652
040	__ a DLC c DLC d YDXCP d BWKUK d BWK d FDA d UKM d BTDTA d CEF d CDX d DLC
050	00 a PR2803.A2 b H35 2009
082	00 a 822.3/3 2 22
100	1_ a Shakespeare, William, d 1564-1616.
245	10 a As you like it / c edited by Michael Hattaway.
250	__ a Updated ed.
260	__ a Cambridge, UK ; a New York : b Cambridge University Press, c 2009.
300	__ a xv, 240 p. : b ill. ; c 24 cm.
490	1_ a New Cambridge Shakespeare
504	__ a Includes bibliographical references (p. 240).
505	0_ a Introduction. Journeys ; Plays within the play ; Theatrical genres ; Pastoral ; Counter-pastoral ; The condition of the country ; Politics ; 'Between you and the women the play must please' ; Gender ; Nuptials ; Sources ; Date and occasion ; Stage history ; Recent critical and stage interpretations -- Note on the text -- List of characters -- The play -- Textual analysis -- Appendixes : 1. An early court performance? ; 2. Extracts from Shakespeare's principal source, Lodge's Rosalind ; 3. The songs.
600	10 a Shakespeare, William, d 1564-1616. t As you like it .
650	_0 a Fathers and daughters v Drama.
650	_0 a Exiles v Drama.
655	_7 a Pastoral drama. 2 gsafd
655	_7 a Comedies. 2 gsafd
700	1_ a Hattaway, Michael.

As you like it [print]



Instance

As you like it [on-line]

Cambridge University Press



Publisher

Shakespeare, William, 1564-1616

Шекспир, У. 1564-1616 Уильям



Agent

Saixpēr, Gouilliam, 1564-1616

As you like it



Work

Come ti piace

Comme il vous plaira

Fathers and daughters



Subject

Padri e figlie

Pères et filles

Cambridge Univ. Press

Data entification, reconciliation, enrichment and publication

Bring together and make available data from different sources in a way that could be defined as *democratic* to better identify the entity in question.

Even wider reconciliation and enrichment processes form the basis of a number of projects that convert and publish bibliographic catalogues as Linked Open Data, such as:

• **Share VDE – *Share Virtual Discovery Environment***. www.share-vde.org (in partnership with Casalini Libri and @Cult)



Albert Camus on the SHARE-VDE platform

The screenshot displays the SHARE-VDE platform interface. At the top, there is a navigation bar with the SHARE logo (a yellow circle with 'SHARE' text) and a network diagram with nodes for 'Virtual', 'Discovery', and 'Environment'. The text 'Person/Work' is written in a script font. There are language options for 'Italiano' and 'Inf', and a 'Go' button. Below the navigation bar is a search bar with a 'Work' icon and a 'Browse' button with a keyboard icon.

The main content area is titled 'Search Work/Publications' and includes 'EXPAND ALL' and 'CLOSE ALL' buttons. On the right, there are icons for RDF, a printer, and an email icon.

The search results are displayed in a central area with a book icon. The main result is 'Galdós and medicine' with ID: 240309. Below it, the 'Creators:' field is empty. To the left, a 'Publications' box lists three entries: 'Galdós and medicine' with a green icon, 'Galdós and medicine' with a blue and red icon, and 'Galdós and Medicine' with a blue and green icon. To the right, a 'Subject' box lists various subjects: 'Pérez Galdós, Benito, 1843-1920', 'Literature and medicine -- Spain -- History -- 19th century', 'Pérez Galdós, Benito, 1843-1920', 'History', 'Spain', 'Medicine in Literature', 'Literature and medicine', '1800-1899', 'History, 19th Century', 'Pérez Galdós, Benito, 1843-1920 -- Criticism and interpretation', 'Criticism, interpretation, etc', '1800 - 1899', 'Pérez Galdós, Benito, 1843-1920 -- Criticism and interpretation', and 'Medicine in literature'.

At the bottom, there is a search bar with a magnifying glass icon and the text '(2424573) Stannard, Michael W., 1942-...'

**A WORK AS AN ENTITY
WITH ITS
RELATIONSHIPS!**

http://share-vde.org/sharevde/searchTitles?t_cluster_id=240309&l=en

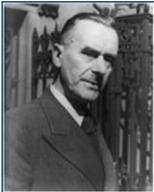
Different entities from the same Marc record!

**HERE THOMAS
MANN IS THE
SUBJECT
OF A WORK!**

The screenshot shows a web interface for a 'Person/Work' database. At the top, there are navigation links for 'Work', 'Publishers', and 'Subject'. A central section features a portrait of Thomas Mann, with the text 'Mann, Thomas, 1875-1955' and 'ID: 215541'. Below this, a vertical line labeled 'Subject in' points to a grid of 24 rounded rectangular buttons, each containing a magnifying glass icon and the title of a work where Thomas Mann is the subject.

Work Publishers Subject

Subject


Mann, Thomas, 1875-1955
ID: 215541
Subject in

 Esmeralda	 "Steine in Hitlers Fenster"	 Lebenstraum und Todesnähe
 Altersgraue Legende"	 Lebens-Werk	 Bürger auf Abwegen
 Schönheit und Verfall	 Faust legend	 Manns
 Literarischer Erfahrungshorizont	 Stationen der Thomas-Mann-Forschung	 Schreibprozess bei Thomas Mann und Franz Kafka und seine didaktischen ...
 Literarische Dekadenz	 Die Musik	 "Ganz entre nous"
 Jahrhundert der Manns	 Kraft der Berührung	 Beslissing
 Wagner und unsere Zeit	 Correspondence	 Lebenstraum und Todesnähe
 King and his rival	 Novel and the Nazi past	

Different entities from the same Marc record!

SHARE Virtual Discovery Environment Person/Work

Work Publishers Subject

Publishers

EXPAND ALL CLOSE ALL

This publisher in

- isni
- WIKIDATA
- LIBRARY OF CONGRESS
- data.bnf.fr
- VZ AF

Wikipedia

Giulio Einaudi editore, nota anche più semplicemente come Einaudi, è una casa editrice italiana.

Wikipedia

Giulio Einaudi editore
ID: 116434

Places

- Torino

Other forms

- Giulio Einaudi Editore casa editrice italiana
- Einaudi, Firm
- Giulio Einaudi editore.
- Giulio Einaudi e editore
- Giulio Einaudi Editore

Person

- Mattei, Ugo.
- Galluccio, Bruno
- Ainis, Michele
- Perniola, Mario, 1941-....
- Lucas, Uliano.
- Marcoaldi, Franco, 1955-...
- Castelnuovo, Enrico, 1929-2014
- La Cecla, Franco, 1950-...
- Tafuri, Manfredo, 1935-1994
- Bre, Silvia
- ...(other)

Subject

- Philosophy, Ancient
- Ariadne (Greek mythology)
- Communication
- Bioethics
- Europe
- World War, 1914-1918
- Rome (Italy)

**THE PUBLISHER
WITH ITS RELATIONSHIPS!**

Entities in *cluster*: an example of collaboration and sharing



Vivaldi, Antonio, 1678-1741
ID: 37154

Questo autore in



Altre forme del nome

- Vivaldi, Antonio, 1678-1741
- VIAF 1678-1741, אנטוניו, אדווינו
- VIAF Vivaldi, Antonio, 1678-1741
- VIAF Vivaldi, Antonio
- VIAF Vivaldi, Antonio, sac., 1678-1741
- VIAF Вивальди, А. 1678-1741 Антонио
- VIAF Вивальди, Антонио, 1678-1741
- VIAF Vivaldi, Antonio, 1680-1741
- VIAF 1741-1678, فيالدي, أنطونيو
- VIAF Antonio Vivaldi compositore e violinista italiano esponente di spicco del tardo barocco veneziano
- VIAF Vivaldi, Antonio, ca.1678-1741
- VIAF Vivaldi, Antonio (Italian composer and musician, 1678-1741)
- Prete rosso, 1678-1741
- Vivaldis, A., 1678-1741
- Vivaldi, A. (Antonio), 1678-1741
- Vivarudi, Antonio, 1678-1741
- Vivaldi, Antonio

...(altre forme)

The result of a reconciliation of the entity *Antonio Vivaldi* in the Share VDE project, with data from different sources and projects:

- the authorized form from a local authority file
- the variant forms originating from the references on the local authority records
- the variant forms originating from the VIAF
- the forms of the name used in the bibliographic records.

The cluster is completed and enriched with identifiers for the same entity, Antonio Vivaldi, from sources such as:

- Wikidata
- Library of Congress Name Authority File
- Data.bnf.fr
- VIAF

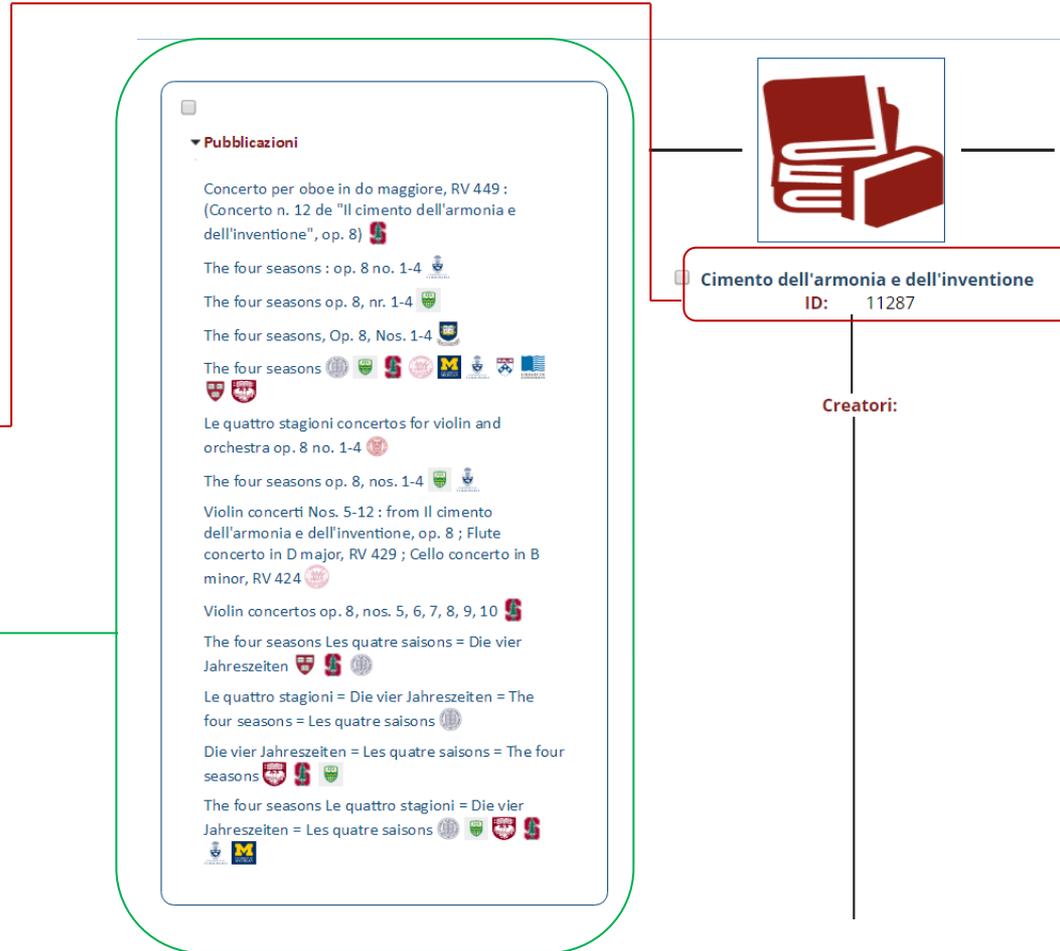


An example of Work/Instances reconciliation

Grouping under a single work title of the many publication titles in the catalogue for *Cimento dell'armonia e dell'invenzione*

Single work title

Brings together different publications/resources present in different catalogues.



http://share-vde.org/sharevde/searchTitles?t_cluster_id=11287



Example of same Instances present in different libraries

The screenshot shows a library catalog interface. At the top, there is a navigation bar with 'SHARE' (a yellow circle), 'Virtual' (a red circle), 'Discovery' (a green circle), and 'Environment' (an orange circle). To the right, there are links for 'Italiano', 'Info', and 'Contacts', and a search bar. The main title is 'Publications(Instances)' in a cursive font. Below the title, there are search filters: 'Advanced search', 'Everywhere', 'show 10 results per page ordered by Score'. On the left, there is a sidebar with 'Info' and 'Library' sections. The 'Library' section lists various institutions like 'University of California Berkeley (2)', 'Columbia University (1)', etc. The main content area shows search results for 'The storm and other things / Eugenio Montale ; translated, with preface and commentary, by William Arrowsmith (9 documenti)'. The results are displayed in a table with columns for 'MARC', 'Tabular view', and 'Short view'. The 'Short view' column shows the following details for each instance:

Library ID	Details
<input type="checkbox"/> ID CORNELL1138554	(Click for details)
<input type="checkbox"/> ID UCBERKELEYUCb125195606	(Click for details)
<input type="checkbox"/> ID UMICH000663201	(Click for details)
<input type="checkbox"/> ID COLUMBIA444093	(Click for details)
<input type="checkbox"/> ID LOC2859531	(Click for details)
<input type="checkbox"/> ID YALE440351	(Click for details)
<input type="checkbox"/> ID HARVARD000456337-9	(Click for details)

At the bottom of the results, it says '5 results found'. On the right side, there is a sidebar with various filters: 'Format' (Language material (13)), 'Bibliographic level' (Monographs (13)), 'Author (Person)' (Montale, Eugenio, 1896-1981 (13), Arrowsmith, William, 1924-1992 (9), Arrowsmith, William, 1924- (4)), 'Author (Corporate)', 'Author (Conference)', 'Works' (Bufera e altro (13)), 'Publication' (Norton (13)), 'Language' (English (13), Italian (3)), and 'Date'.



Reconciliation & Enrichment Automatic procedures



How reconciliation is obtained

Data reconciliation and enrichment is obtained by:

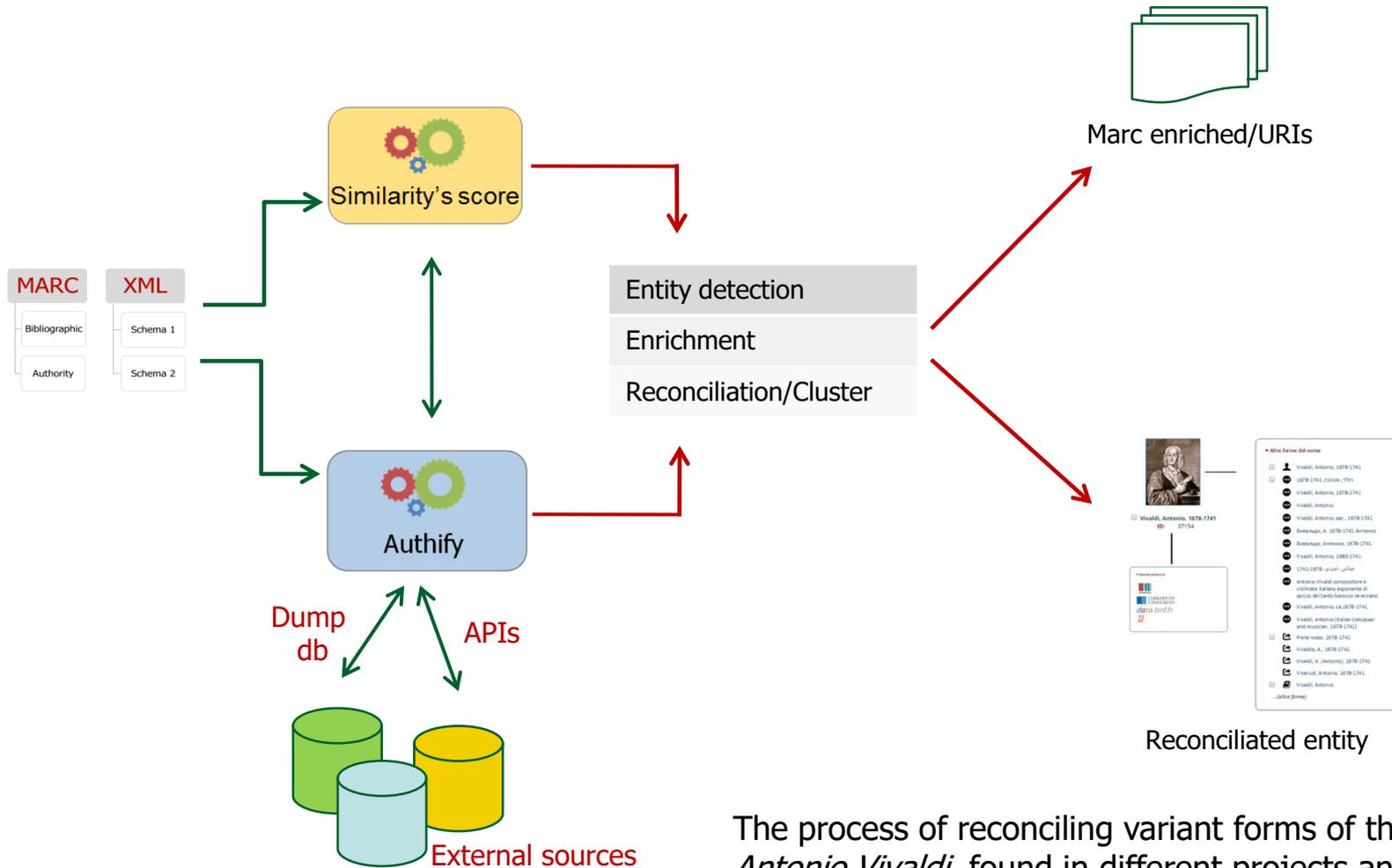
- *automated processes*
- *manual processes*

It is important to underline how the **relationship between the reconciliation and validation of the results** can differ greatly between the automated and manual processes:

- automated processes: a high level of reconciliation and clustering; a low level of result validation;
- manual processes: a low level of reconciliation and clustering; a high level of result validation.



Automated reconciliation and enrichment



The process of reconciling variant forms of the entity *Antonio Vivaldi* found in different projects and catalogues.

Authify – General description

Authify is a RESTful module that offers several search and detection services. The original aim of the project was to overcome some of the limitations of the public VIAF Web API.

VIAF, being a public project, does not allow a massive invocation of its API: for those use cases where such requirement is needed, the project provides a download of the whole dataset.

That was the main reason we started implementing Authify: to index and store the VIAF clusters dataset and also provide powerful full-text and bibliographic search services.

It is possible to add to Authify other dump databases, coming by external projects that make them available.



Authify – Cluster search services

The Authify cluster search service provides, as the name suggests, a full-text search service for names and works clusters. The search Web API uses an “invisible queries” approach in order to (try to) find as precise as possible a match within the managed clusters.

The invisible queries approach makes everything transparent to the caller: in addition to a single search request, the system carries out a chain of different search strategies with different priorities, and the first match to produce a result will populate the response that will be returned.

For debugging purposes, the response will also include the matching strategy that produced the results.



Authify – Cluster search services

The system has been built with extensibility in mind, so the chain is fully configurable; for instance, here is a brief description of the current configuration when searching names clusters:

- **Subfields matching**: the query language allows the caller to specify the source tag / subfields that compose the heading (which is the actual input query string).
- **Input heading exact match**: the system tries to find an exact match with the provided query string.
- **FullText search**: if an exact match is not possible, then a regular full text search is carried out, with options like proximity search for names (e.g. Bertrand Meyer = Meyer Bertrand) and special detection for some entities (e.g. birth and death dates).
- Finally, the system executes a **search by “initials”**, in order to find a valid match in those cases when the input string (or the indexed heading) contains the name in its short form. As with the previous point, this could lead to a less precise response.



Authify – Cluster search services - Response

The query interface: http://labs.atcult.it/authify/names?q=bertrand_Meyer : the system will provide a response like this:

```
{
  "responseHeader" : {
    "QTime" : 3,
    "matching-strategy" : "name::headings-exact-match",
    "status" : 0
  },
  "response" : {
    "docs" : [ {
      "id" : "51714577",
      "type" : "Personal",
      "uri" : "http://viaf.org/viaf/51714577/",
      "headings" : [
        "Meyer, Bertrand, 1950-....",
        "Bertrand Meyer",
        "Meyer, Bertrand" ],
      "sources" : [
        "BNF|12079479",
        "DNB|112127843",
        "ISNI|0000000109003927",
        "LC|n 86061235",
        "LNB|LNC10-000142119",
        "NDL|00471567",
        "NKC|skuk0004073",
        "NLA|000035194108",
```



Authify – Relator term detection

Another service which has been added to Authify is the so called “**Relator term detection**”.

Starting from a MARC record (whatever the specific dialect) the system analyses all (configured) tags that contain a name and, for each of them, tries to determine (using the statements of responsibility of the input record) what is the corresponding role within the work represented by the given record.

So, for instance, for the following input (the example shows only the relevant tags):

245 10\$aFondamenti di teoria dei circuiti /\$cCharles A. Desoer, Ernest S. Kuh ;
prefazione all'edizione italiana di G. Biorci

100 1 \$aDesoer, Charles A.

700 1 \$aBiorci, Giuseppe

700 1 \$aKuh, Ernest S.



Authify – Relator term detection

The system will give a response like this:

```
{
  "id": "LE02614324",
  "statements": [
    "245 10$aFondamenti di teoria dei circuiti /$cCharles A. Desoer, Ernest S. Kuh ; prefazione all'edizione italiana di G. Biorci"
  ],
  "names": [
    "100 1 $aDesoer, Charles A.",
    "700 1 $aBiorci, Giuseppe",
    "700 1 $aKuh, Ernest S."
  ],
  "responsibilities": {
    "content": {
      "http://id.loc.gov/vocabulary/relators/oth": {
        "headings": [
          {
            "name": "Biorci, Giuseppe"
          }
        ],
        "relatorTermCode": "oth",
        "relatorTermText": "Other"
      },
      "http://id.loc.gov/vocabulary/relators/aut": {
        "headings": [
          {
            "name": "Kuh, Ernest S."
          },
          {
            "name": "Desoer, Charles A."
          }
        ],
        "relatorTermCode": "aut",
        "relatorTermText": "Author"
      }
    }
  }
}
```



Authify – Relator term detection

In these examples you can see that two main roles have been detected:

- **authors**
- **other** (unclassified role).

The “other” role is a catch-all role used when no valuable information can be gathered from the analysis.

Behind a simple token matching analysis, there is a more complicated logic that tries (using, among other things, the search services described in the previous point) to find the role of each name using its variant forms or using a set of tokens that could identify it (e.g. edited by, by, illustrated by).



Entity detection (example 1)

=LDR 00833nam a2200217 i 4500

=001 LE02519084

=005 20020503192020.0

=008 970703s1990\\uk\\eng\\

=020 \\a0415030889

=040 \\aFac. Economia\$bita

=082 0\$a820.9

=100 1\$aStephens, John

**=245 10\$aLiterature, language and change :\$bfrom Chaucer to the present /\$cJohn
Stephens and Ruth Waterhouse**

=260 \\bRoutledge,\$cc1990

=300 \\aix, 293 p. ;\$c20 cm.

=650 \4\$aLetteratura inglese\$xStoria e critica

=650 \4\$aLingua inglese

=700 1\$aWaterhouse, Ruth



Entity detection - Authify/Detect response (1)

Response Body service authify/detect:

```
{
  "id": "LE02519084",
  "statements": [
    "245 10$aLiterature, language and change :$bfrom Chaucer to the present /$cJohn Stephens and Ruth Waterhouse"
  ],
  "names": [
    "100 1 $aStephens, John",
    "700 1 $aWaterhouse, Ruth"
  ],
  "responsibilities": {
    "content": {
      "http://id.loc.gov/vocabulary/relators/aut": {
        "headings": [
          {
            "name": "Stephens, John"
          },
          {
            "name": "Waterhouse, Ruth"
          }
        ],
        "relatorTermCode": "aut",
        "relatorTermText": "Author"
      }
    }
  }
}
```



Entity detection (example 2)

=LDR 01127pam a2200325 a 4500

=001 7486885

=005 20150720142401.0

=008 090901t20152015mauab\\\\b\\\\001\0\eng\\

=010 \\\$a 2009036444

=020 \\\$a9781566567879\$qpaperback

=020 \\\$a1566567874\$qpaperback

=024 \\\$a99963025763

=035 \\\$a(OCOLC)908588988

=035 \\\$a(OCOLC)ocn908588988

=035 \\\$a(NNC)7486885

=040 \\\$aDLC\$beng\$cDLC\$dBTCTA\$dBDX\$dOCLCF\$dOCLCO\$dMNM\$dNhCcYBP

=043 \\\$aa-is---\$aawba---

=050 00\$aDS109.93\$b.J48 2015

=082 00\$a956.94/4205\$222

=245 00\$aJerusalem interrupted :\$bmodernity and colonial transformation 1917-present /\$cedited and introduced by Lena Jayyusi.

=260 \\\$aNorthampton, Mass. :\$bOlive Branch Press,\$c2015.

=300 \\\$axxii, 499 p. :\$bill., maps ;\$c24 cm.

=504 \\\$aIncludes bibliographical references and index.

=651 \0\$aJerusalem\$xHistory\$y20th century.

=651 \0\$aJerusalem\$xHistory\$y21st century.

=651 \0\$aJerusalem\$xInternational status.

=650 \0\$aArab-Israeli conflict.

=700 1\\\$aJayyusi, Lena.



Entity detection - Authify/Detect response (2)

```
{
  "id": "7486885",
  "statements": [
    "245 00$aJerusalem interrupted :$bmodernity and colonial transformation 1917-present /$cedited and
    introduced by Lena Jayyusi."
  ],
  "names": [
    "700 1 $aJayyusi, Lena."
  ],
  "responsibilities": {
    "content": {
      "http://id.loc.gov/vocabulary/relators/edt": {
        "headings": [
          {
            "name": "Jayyusi, Lena."
          }
        ],
        "relatorTermCode": "edt",
        "relatorTermText": "Editor"
      }
    }
  }
}
```



Entity detection (example 3) - Critical case

=LDR 01145nam a2200241 i 4500

=001 LE01988135

=005 20020503105244.0

=008 010702s1999\\it\\000\0\lat\\

=020 \\\$a882092868X

=040 \\\$aDip.to Beni Arti e Storia\$bita

=082 0\$a264.024

=245 00\$aBreviarium Romanum :\$beditio princeps, 1568 /\$cedizione anastatica, introduzione e appendice a cura di Manlio Sodi, Achille Maria Triacca ; con la collaborazione di Maria Gabriella Foti ; presentazione di Virgilio Noè

=260 \\\$aCittà del Vaticano :\$bLibreria editrice Vaticana,\$c1999

=300 \\\$aXXII, 1056 p. ;\$c25 cm

=440 \0\$aMonumenta liturgica concilii tridentini\$v3

=700 1\\\$aSodi, Manlio

=700 1\\\$aTriacca, Achille Maria

=700 1\\\$aFoti, Maria Gabriella

=700 1\\\$aNoè, Virgilio

=907 \\\$a.b10000914\$b02-04-14\$c29-05-02



Entity detection - Authify/Detect response (3)

```
{
  "id": "LE01988135",
  "statements": [
    "245 00$aBreviarium Romanum :$beditio princeps, 1568 /$cedizione anastatica, introduzione e appendice a cura di Manlio Sodi, Achille Maria Triacca ; con la collaborazione di Maria Gabriella Foti ; presentazione di Virgilio Noè"
  ],
  "names": [
    "700 1 $aFoti, Maria Gabriella",
    "700 1 $aNoè, Virgilio",
    "700 1 $aSodi, Manlio",
    "700 1 $aTriacca, Achille Maria"
  ],
  "responsibilities": {
    "content": {
      "http://id.loc.gov/vocabulary/relators/oth": {
        "headings": [
          {
            "name": "Sodi, Manlio"
          },
          {
            "name": "Triacca, Achille Maria"
          },
          {
            "name": "Foti, Maria Gabriella"
          },
          {
            "name": "Noè, Virgilio"
          }
        ],
        "relatorTermCode": "oth",
        "relatorTermText": "Other"
      }
    }
  }
}
```



Name cluster process

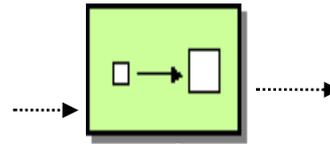
Authority form:

Lucio, José de

De Lucio, José

Lucio, J. de (José de)

Lucio, José de



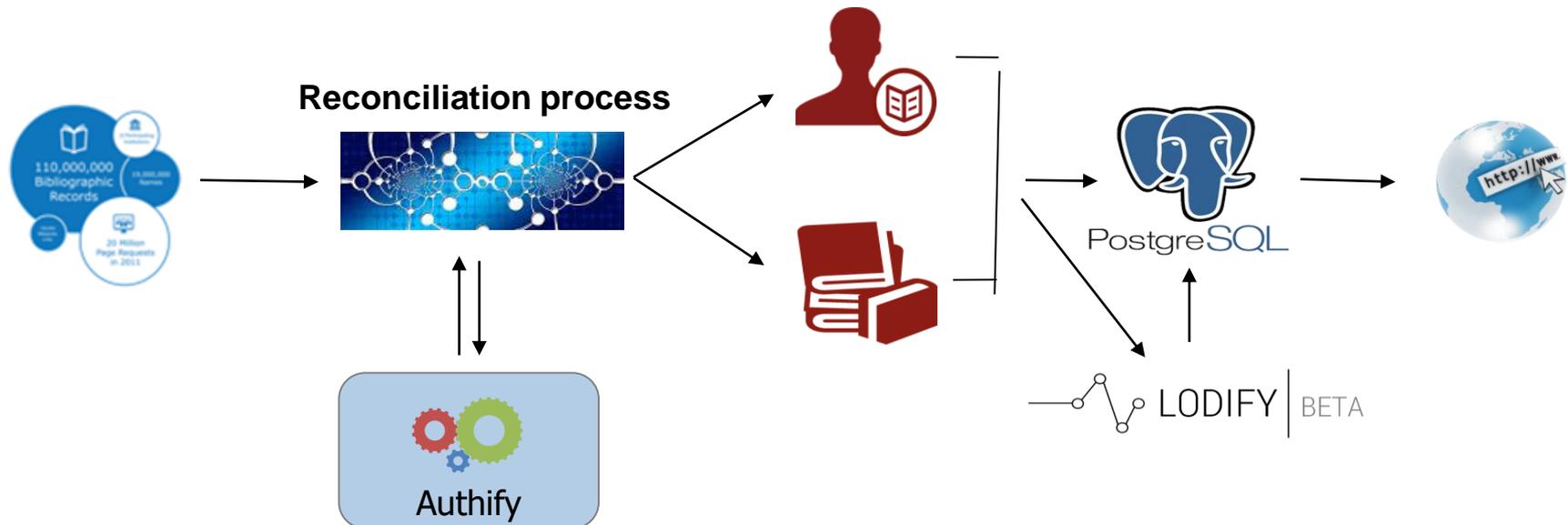
ID cluster: 2085026
Author : Lucio, José de m. 1949

Other forms:

Lucio, José de
Lucio, José de m. 1949
De Lucio, José
Lucio, J. de (José de)

Massive clusters processes

- Authority headings analysis and process in PostgreSQL;
- Data enrichment with external sources
- MARC bibliographic process
- Entity detection (authors and co-authors identification process)
- Name heading-to-Authority names association (through a weighted comparison algorithm)
- Name heading-to-Variant names association
- Cluster check (it exists = add; it doesn't exist = create new)



Reconciliation & Enrichment Manual procedures



PCC directives

PCC identifies and addresses policy issues on the use of identifiers in MARC:



developing guidelines to include identifiers in MARC bibliographic and authority records



the use of multiple identifiers for the same entity



determining the entities for which identifiers should be provided in an initial implementation



identifying automated methods for populating and maintaining new and existing records with identifiers

The importance of identification and detection in the Semantic Web



Key elements of the cataloguing workflow:
entity identification
reconciliation

To enrich a MARC record with URIs Casalini Libri uses the “URI MANAGEMENT SYSTEM” (included in the OLISuite cataloguing module within WeCat).

This also simplifies the reconciliation of varying forms of the same entity with the development of detection procedures for entity identification and the conversion to BIBFRAME.

The manual process to enrich MARC records

The “URI MANAGEMENT SYSTEM” allows the management of **multiple identifiers** for each **access point** or **heading**.

Use of external sources (such as NAF, ISNI and VIAF) with API and web services.

Associate heading with the URIs that identify it in each of the projects.



URI Management System in OLISuite

Search/Cataloguing

- Simple
- Advanced
- External
- Last index
- Active queries
- Indexes list
- Diacritic

Reports

- General
- Label

Templates/Download

From template

- Load from file
- Check loaded data
- Templates
- Load from VIAF

Access

- New heading (F9)
- Transfer relationships

Browse search

Any AMICUS Database

Advanced >> Browse search

kafka in None and display 10 terms **Scan for**

Configuration

NT References Aut. Doc. Level Indic. Acc. Rapid insertion 

Heading	DB	NT	URI		Refs	A	Docs	Level	Index.	Acc.
 Kaestli, Jean-Daniel	B1	0	0	<input type="checkbox"/>	0	0	5	Unverified	und	und
 Kafka, Franz, 1883-1924	B1	1	2	<input type="checkbox"/>	0	0	1	Saved	und	und
 Kafka, Franz 1883-1924 The Metamorphosis	B1	0	0	<input type="checkbox"/>	0	1	0	Saved	und	eng
 Kaftal, George	B1	0	0	<input type="checkbox"/>	0	0	4	Unverified	und	und

URI Management System in OLISuite

The screenshot displays the 'URI management' interface. On the left, there are three main sections: 'Search/Cataloguing' with options like Simple, Advanced, External, Last index, Active queries, Indexes list, and Diacritic; 'Reports' with General and Label; and 'Templates/Download' with a dropdown menu set to 'From template' and options for Load from file, Check loaded data, Templates, and Load from VIAF. The main content area is titled 'URI management' and includes a dropdown for 'Any AMICUS Database'. Below this, a breadcrumb trail reads 'Advanced >> Browse search >> URI management'. The central focus is a light blue box for the 'Selected heading: Kafka, Franz, 1883-1924', which contains a table of identifiers. The table has columns for Source, Http Uri, a checkbox, Validated status, and Options. Two entries are shown: NAF and ISNI, both with 'Validated' status marked as 'X'. A 'New' button is at the top right and a 'Delete' button is at the bottom right of the table area. A red arrow points from the table area to a callout box at the bottom.

URI management Any AMICUS Database

Advanced >> Browse search >> URI management

Selected heading: Kafka, Franz, 1883-1924 New

Source	Http Uri	<input type="checkbox"/>	Validated	Options
NAF	http://id.loc.gov/authorities/names/n81063091	<input type="checkbox"/>	X	
ISNI	http://isni.org/isni/000000012280370X	<input type="checkbox"/>	X	

Delete

The cataloguer can check, modify, delete or add other identifiers to the same heading

Adding new URIs to a heading in OLISuite

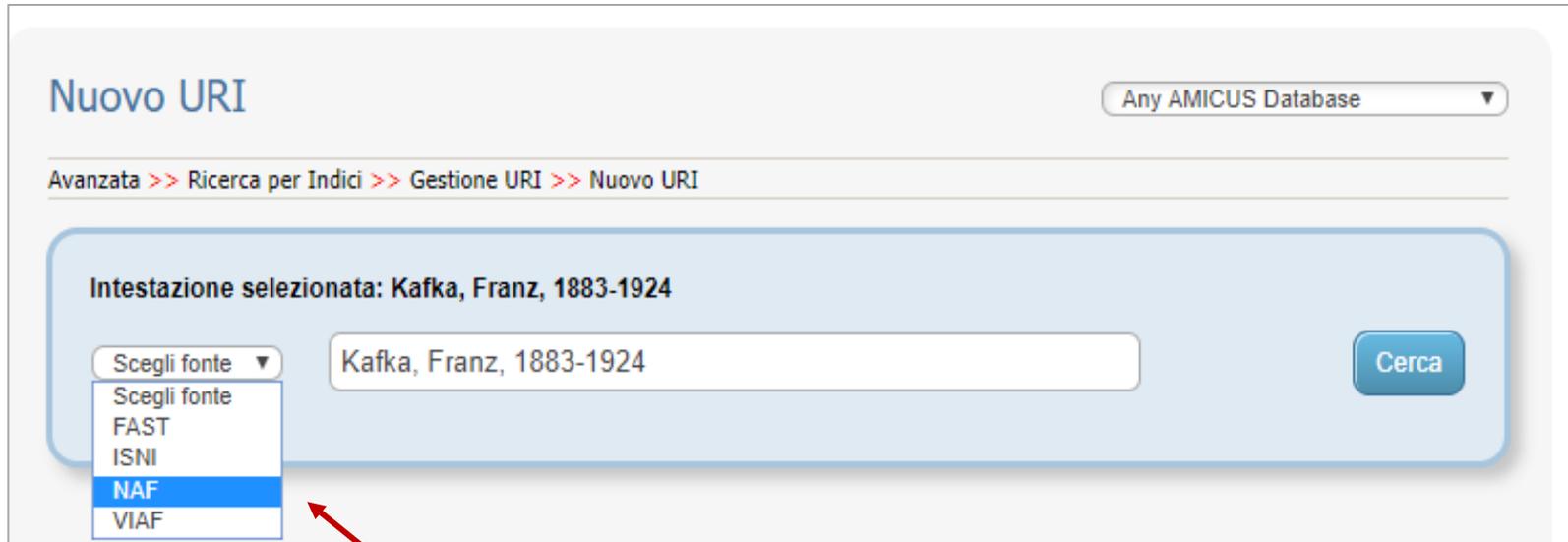
Nuovo URI Any AMICUS Database ▾

Avanzata >> Ricerca per Indici >> Gestione URI >> Nuovo URI

Intestazione selezionata: Kafka, Franz, 1883-1924

Scegli fonte ▾ Cerca

- Scegli fonte
- FAST
- ISNI
- NAF**
- VIAF



From this drop down menu the cataloguer can choose the desired source and start the URI search

Adding new URI to a heading OLISuite

Nuovo URI Any AMICUS Database ▾

Avanzata >> Ricerca per Indici >> Gestione URI >> Nuovo URI

Intestazione selezionata: Kafka, Franz, 1883-1924

NAF ▾ Cerca

Salva

Fonte	Link	Etichetta	Aggiungi
NAF	http://id.loc.gov/authorities/names/n81063091	Kafka, Franz, 1883-1924	<input checked="" type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/n82140681	Kafka, Franz, 1883-1924. Short stories. Selections. English	<input type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/n82125876	Kafka, Franz, 1883-1924. Short stories. English	<input type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/no2005006616	Kafka, Franz, 1883-1924. Brief an den Vater. English	<input type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/no98037666	Kafka, Franz, 1883-1924. Works. 1995	<input type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/n2012039411	Kafka, Franz, 1883-1924. Briefe an Milena. Hebrew	<input type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/n88114203	Kafka, Franz, 1883-1924. Erzählungen und kleine Prosa. English	<input type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/n91003020	Kafka, Franz, 1883-1924. Diaries	<input type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/nr97033862	Kafka, Franz, 1883-1924. Works. 1946	<input type="checkbox"/>
NAF	http://id.loc.gov/authorities/names/nr99034045	Kafka, Franz, 1883-1924. Amerika. English	<input type="checkbox"/>

From the search result window choose the desired URI and SAVE

Access points and URIs

The multiple identifiers associated with Kafka are saved in a specific oracle table and not directly in subfield \$0 of the MARC tag for that heading.

Fonte	Http Uri	 Validato	Opzioni
Local	http://uri.casalini.it/names/75427	<input type="checkbox"/>	 
ISNI	http://isni.org/isni/000000012280370X	<input type="checkbox"/>	   
VIAF	http://viaf.org/viaf/56611857/	<input type="checkbox"/>	   

While it is an acceptable practice in MARC to have multiple identifiers for the same entity in one field via repeating subfields, that does not translate well to RDF.

It is impossible for the program to determine which subfield each \$0 URIs references because the sequence and order of subfields has no meaning for the program.

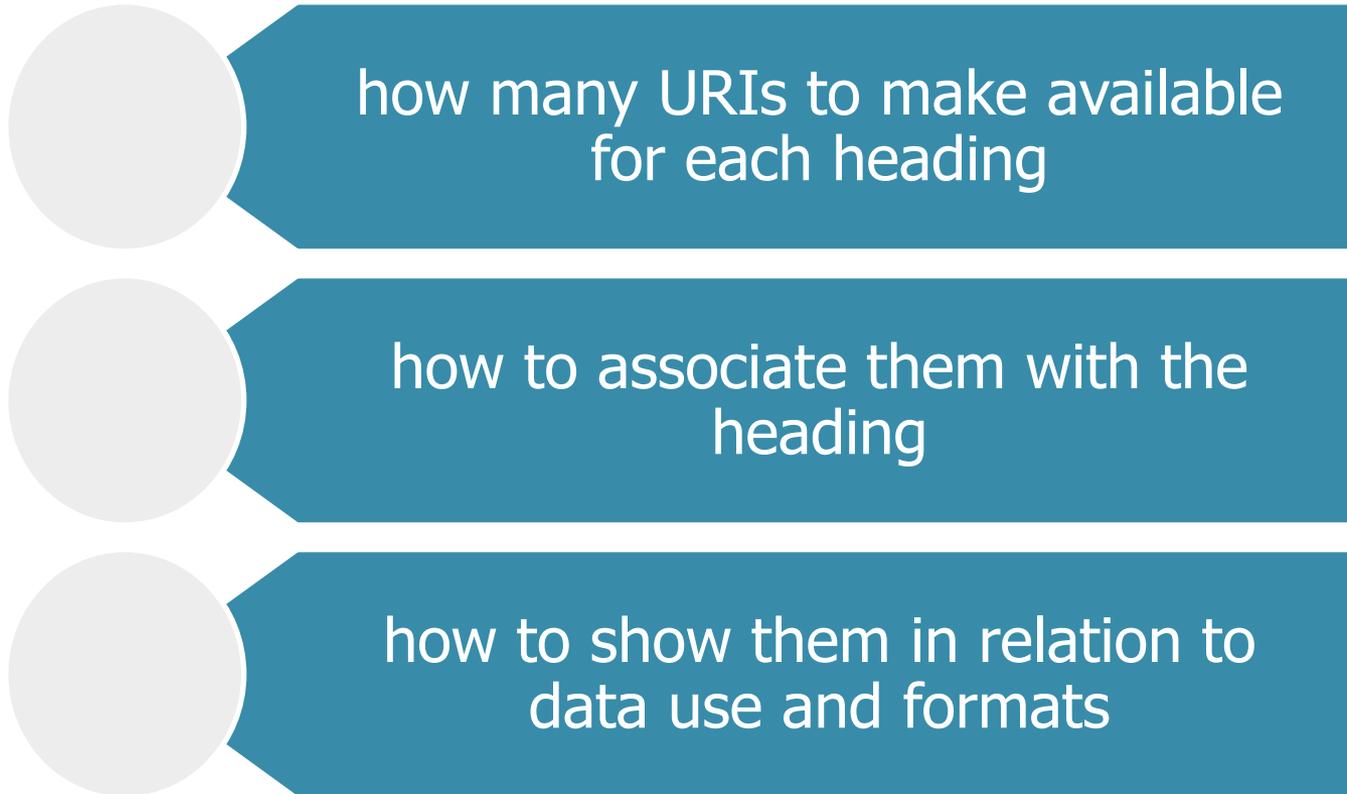
For example:

```
382 0\${aviolin$n1$n1$s2$2lcmpt  
$0http://id.loc.gov/authorities/performanceMediums/mp2013015782 ${apiano  
$0http://id.loc.gov/authorities/performanceMediums/mp2013015550
```



Access points and URIs

Saving the different URIs in an Oracle table allows them to be used in various ways, selected during the data export/conversion:



Different customer profiles that were previously defined in Adempiere are considered.

Adempiere – Customer profile

Anagrafica Contratti Clienti SuperUser@Casalini Libri Spa.

File Modifica Visualizza Vai Strumenti Finestra Aiuto

Dati Globali

Validità

Approval/NTI

Righe Contratto

Dati Bibliografici

Approval/NTI Dettaglio

Materie

Suddivisioni Temporali

Suddivisioni Geografiche

Suddivisioni Particolari

Lingua/Paese

LC

Progetto MAG0DABN Manuale

Business Partner MAG0_000_Harvard University

Ente **MAG0DA-Harvard College Library - Monograph Acquisitions European Team**

Codice Famiglia GFAM006

Numero Contratto MAG0DABN

Tipologia contratto NTL_New Title Information

Descrizione

Mappatura MARC M0000106_Mappatura - MAG0DA tutti

Mappatura URI **MU000001**

Appalto/Vendor Code

Note Cliente AS

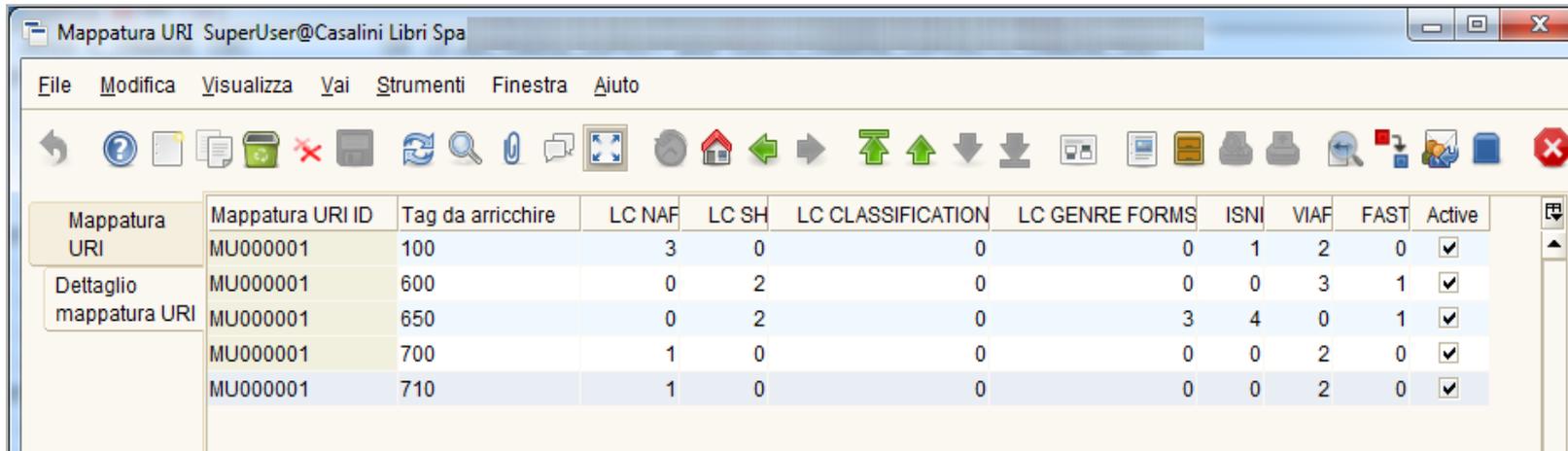
Stato contratto ATT_Activo

**Customer profile for
Harvard College Library**



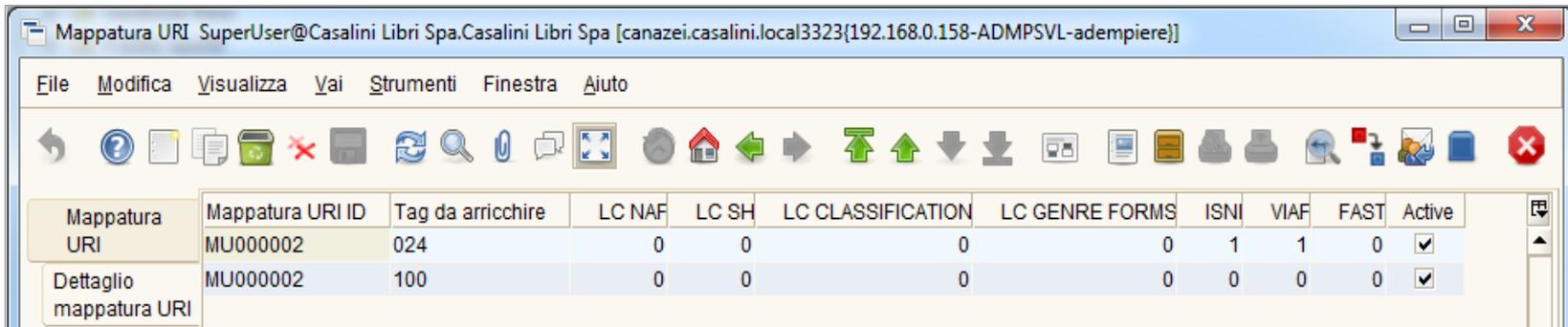
URIs/tag mapping

URIs/tag mapping for Library A



Mappatura URI	Mappatura URI ID	Tag da arricchire	LC NAF	LC SH	LC CLASSIFICATION	LC GENRE FORMS	ISNI	VIAF	FAST	Active
Mappatura URI	MU000001	100	3	0	0	0	1	2	0	<input checked="" type="checkbox"/>
Dettaglio mappatura URI	MU000001	600	0	2	0	0	0	3	1	<input checked="" type="checkbox"/>
	MU000001	650	0	2	0	3	4	0	1	<input checked="" type="checkbox"/>
	MU000001	700	1	0	0	0	0	2	0	<input checked="" type="checkbox"/>
	MU000001	710	1	0	0	0	0	2	0	<input checked="" type="checkbox"/>

URIs/tag mapping for Library B



Mappatura URI	Mappatura URI ID	Tag da arricchire	LC NAF	LC SH	LC CLASSIFICATION	LC GENRE FORMS	ISNI	VIAF	FAST	Active
Mappatura URI	MU000002	024	0	0	0	0	1	1	0	<input checked="" type="checkbox"/>
Dettaglio mappatura URI	MU000002	100	0	0	0	0	0	0	0	<input checked="" type="checkbox"/>



Access point and URIs (customer A)

As \$0 or \$1 is associated to the access point in the MARC bibliographic record:

```
=LDR 00560nam a2200181 4500
=001 000000127573
=003 CaOOAMICUS
=005 20160108094931.0
=008 160107s\\|||||it|||||||000\u\ita\r
=040 \\$aAtCult$bita
=100 1\\$aKafka, Franz,$d1883-
1924$0//isni.org/isni/000000012280370X
=245 03$aLa metamorfosi /$cFranz Kafka.
=260 \\$aMilano :$bLa spiga,$c2002.
=300 \\$a61 p.; $c18 cm
=336 \\$atext$2rdacontent
=337 \\$aunmediated$2rdamedia
=338 \\$avolume$2rdacarrier
=997 \\$aPS
```



URI history => URI Registry

The reorganisation of a cluster can modify its original content, so we need to save the relevant cluster updates in a URI Registry.

The URI Registry could keep information such as (but not limited to):

- the resources added to the cluster, but also modified or removed from it
- the date of the update
- the particular operation performed
- the status of an URI (for instance valid or invalid)
- the URI aliases

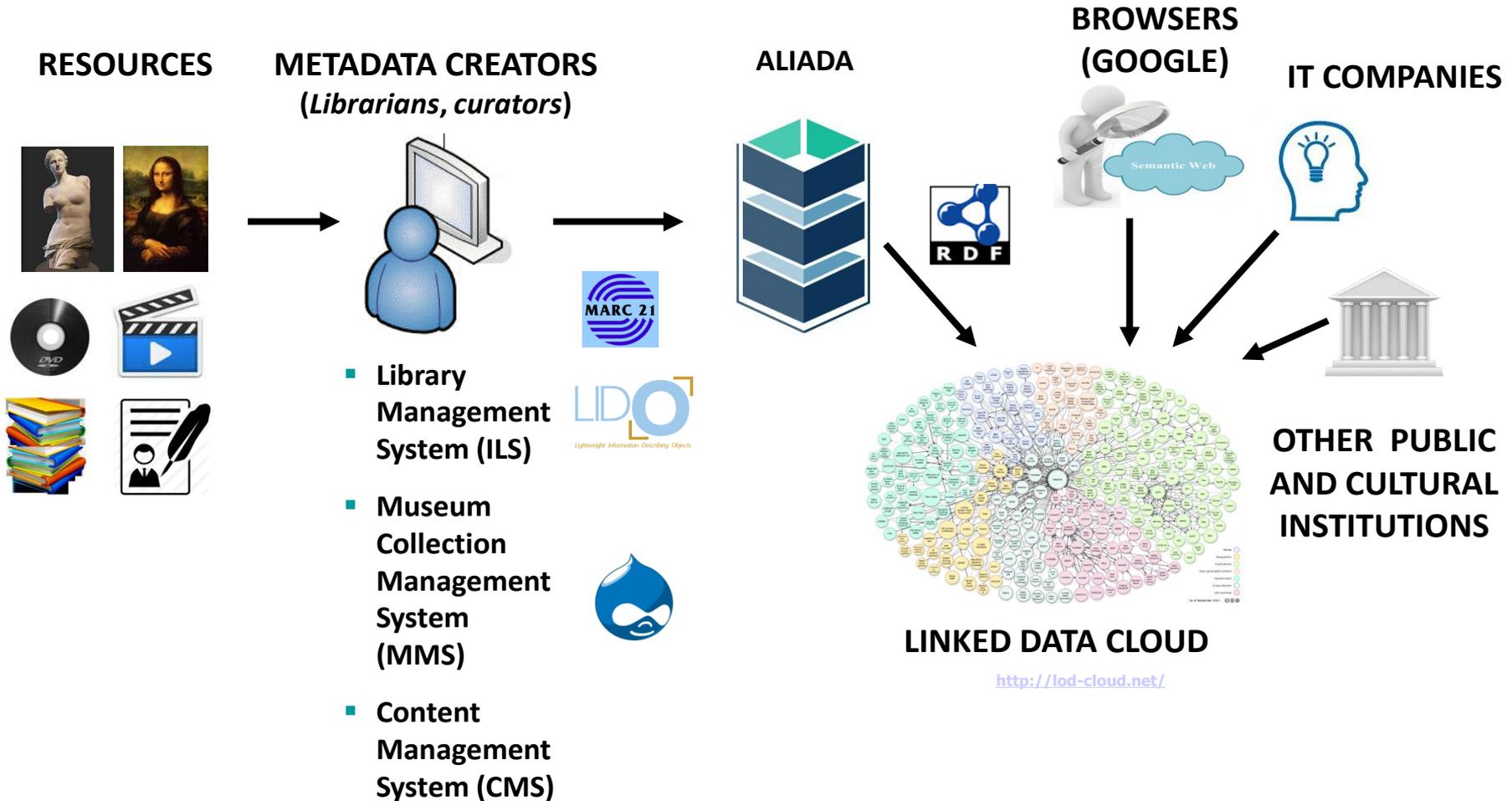


Conversion in RDF/BIBFRAME



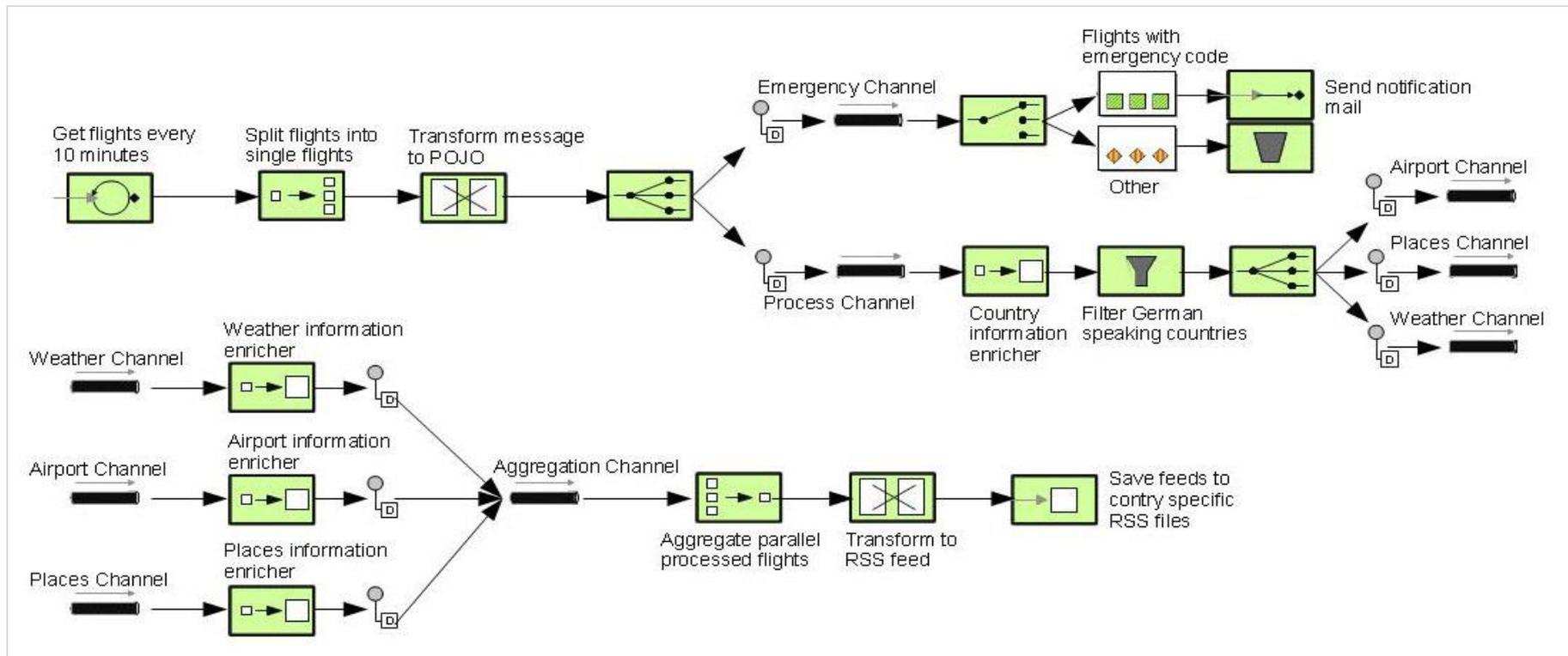
Lodify: the evolution of Aliada for BIBFRAME conversion

The conversion process from any format to RDF



Lodify - The asynchronous pipeline

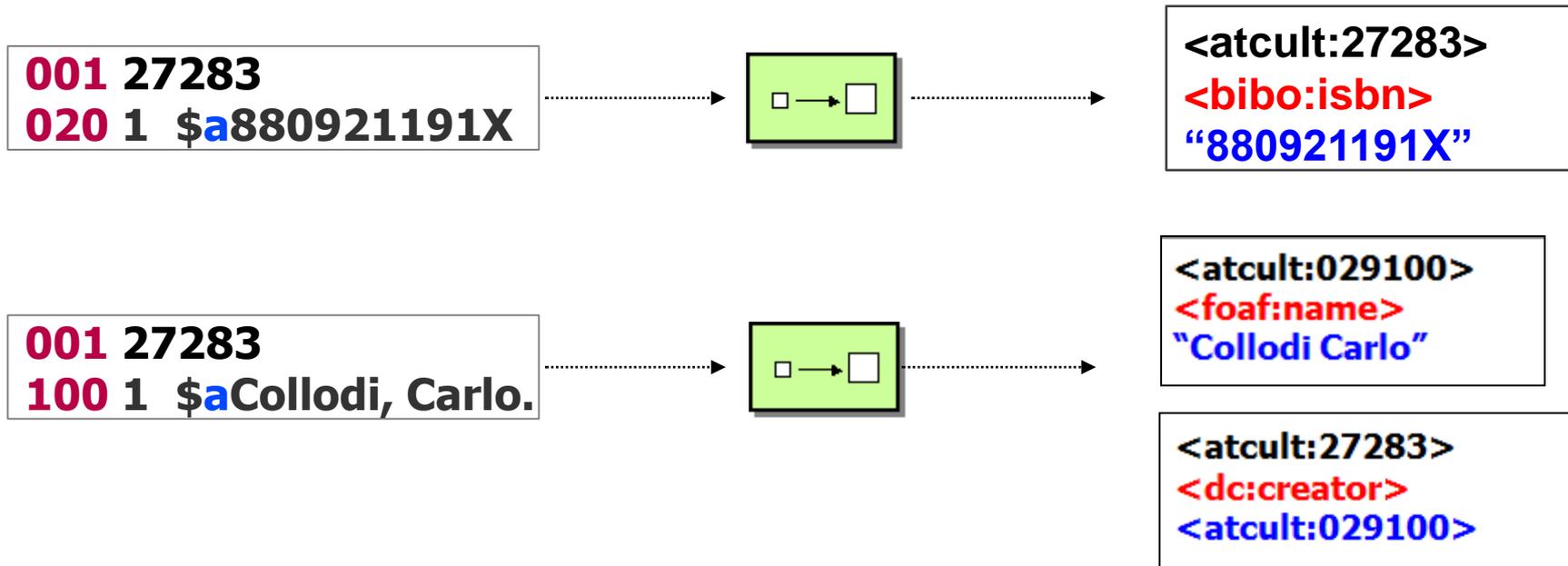
A Lodify building block, realized through Apache Camel. The process is split into atomic pieces (processors), each of these responsible for a small part of the overall task. Each processor can act as a splitter or aggregator and can achieve content manipulation on the incoming message.



Lodify - Conversion templates

Lodify converts each incoming record by means of Conversion templates. Each template associates:

- a MARC record belonging to the incoming data-stream
- with a set of (conversion) rules associated with one or more ontologies.



Trust and Provenance



Guarantee of authority and quality in the new LOD environment

Need to guarantee the accuracy of this information

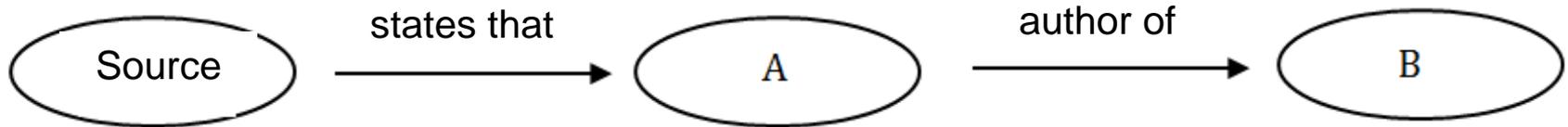
Knowing the *provenance* of a piece of information – *its origin*, authorship or matrix – is a key factor in determining *the extent to which it can be trusted*.

The information source has become the guarantor of quality: creating a link between information and its source has become essential for the purpose of guaranteeing the authority of the information itself.



Guarantee of authority and quality in the new LOD environment

The source or *provenance*, which, in turn, must be constructed with reference to specific ontologies, providing the classes, properties and restrictions needed for identifying it, becomes the *fourth element* added to every triple (assertion) to certify its validity, transforming the triple into a quadruple.



Stating the *provenance* of a piece of information is an essential element for increasing the trust that can be placed in data, and facilitating its use and sharing by end users or by the institutions choosing to co-operate in this way.



SHARE-VDE phase 2 deliverables



Phase 2 deliverables overview



BIB1

Deliverable 1: The datasets in BIBFRAME 2.0 of the entire catalogue of each institution with the "tuples" derived directly from MARC records, delivered both as triples and as quadruples with the addition of provenance and with Share-VDE URIs



BIB2



BIB...



Clusters Knowledge base

Deliverable 2: The knowledge base of clusters accessible in RDF



BIB1



BIB2



BIB...

Deliverable 3: The datasets in BIBFRAME 2.0 for each institution with the triples that include the URIs from the external sources

External (VIAF) URIs

SHARE-VDE URIs



Marc enriched
(Binary)
(one for LIB)

Deliverable 4: The MARC21 records for each institution enriched with URIs



Phase 2 deliverables overview

Linked

D1

the catalogue of each library converted into BIBFRAME 2.0 format

Entities are reconciled in the dataset and linked to SHARE-VDE project URIs of D2 for identification.

D2

the SHARE-VDE project Knowledge base of clusters in RDF format

Common for all institutions as it includes data from all of the participants. Entities in D2 are enriched with URIs from external sources. All variant forms are included.

D3

the dataset converted in BIBFRAME 2.0 with external URIs included

This dataset includes a certain number of relationships already present in the knowledge base. Works autonomously from D2.

D4

the MARC21 version of D3

Includes all of the institution's records enriched with URIs.



BF LC vocabulary extensions: The current situation

Class List

<u>AppliesTo</u>	<u>DemographicGroup</u>	<u>Relation</u>
<u>CreatorCharacteristic</u>	<u>EncodingLevel</u>	<u>Relationship</u>

Property List

<u>applicableInstitution</u>		
<u>appliesTo</u>		
<u>creatorCharacteristic</u>		
<u>demographicGroup</u>		
<u>encodingLevel</u>		
<u>metadataLicensor</u>	<u>name10MatchKey</u>	<u>title00MarcKey</u>
<u>name00MarcKey</u>	<u>name11MarcKey</u>	<u>title00MatchKey</u>
<u>name00MatchKey</u>	<u>name11MatchKey</u>	<u>title10MarcKey</u>
<u>name10MarcKey</u>	<u>primaryContributorName00MatchKey</u>	<u>title10MatchKey</u>
	<u>primaryContributorName10MatchKey</u>	<u>title11MarcKey</u>
	<u>primaryContributorName11MatchKey</u>	<u>title11MatchKey</u>
	<u>projectedProvisionDate</u>	<u>title30MarcKey</u>
	<u>relation</u>	<u>title30MatchKey</u>
	<u>relationship</u>	<u>title40MarcKey</u>
		<u>title40MatchKey</u>
		<u>titleSortKey</u>



Deliverable 1: the RDF conversion (section of RDF conversion)

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ns0: <http://id.loc.gov/ontologies/bflc/> .
@prefix ns1: <http://id.loc.gov/ontologies/bibframe/> .
@prefix ns2: <http://id.loc.gov/vocabulary/relators/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ns3: <http://www.loc.gov/mads/rdf/v1#> .
<http://share-vde.org/sharevde/rdfBibframe2/Agent/LOC> rdfs:label "LOC" .
<http://id.loc.gov/authorities/names/n50044402>
  a <http://id.loc.gov/ontologies/bibframe/Agent>, <http://id.loc.gov/ontologies/bibframe/Person> ;
  rdfs:label "Bridget, approximately 1303-1373." ;
  ns0:name00MatchKey "Bridget, approximately 1303-1373." ;
  ns0:name00MarcKey "1000 $aBridget, $cof Sweden, Saint, $dapproximately 1303-1373." ;
  $0http://id.loc.gov/authorities/names/n50044402" ;
  ns0:primaryContributorName00MatchKey "Bridget, approximately 1303-1373." .
<http://id.loc.gov/authorities/names/n84007202>
  a <http://id.loc.gov/ontologies/bibframe/Work>, <http://id.loc.gov/ontologies/bibframe/Text> ;
  ns1:title <http://share-vde.org/sharevde/rdfBibframe2/Title/a0f4e860-c259-3611-b7ca-03b9369568cb>, <http://share-
vde.org/sharevde/rdfBibframe2/uniform-title/a0f4e860-c259-3611-b7ca-03b9369568cb> ;
  ns2:cre <http://id.loc.gov/authorities/names/n50044402> ;
  ns1:content <http://rdaregistry.info/termList/RDAContentType/1020>, <http://id.loc.gov/vocabulary/contentTypes/txt
;
  ns1:adminMetadata <http://share-vde.org/sharevde/rdfBibframe2/AdminMetadata/4a8a08f0-9d37-3737-9564-9038408b5f33>
  ns1:language <http://id.loc.gov/vocabulary/languages/eng>, <http://share-
vde.org/sharevde/rdfBibframe2/Language/74e6a8b1-11ea-3da1-a7d0-a596f4c35208> ;
  ns1:genreForm <http://id.loc.gov/vocabulary/marcgt/bib> ;
  ns1:classification <http://share-vde.org/sharevde/rdfBibframe2/Lcc/842e73eb-0d86-374b-a4bd-dae01b30c68a> ;
  ns1:subject <http://share-vde.org/sharevde/rdfBibframe2/Subject/6c09b127-ac1a-3428-ad6e-f896dbf69260>,
<http://share-vde.org/sharevde/rdfBibframe2/Subject/2f1b63c0-98ed-35e9-b57b-f7d7ce7c525d> .
```

LC vocabulary extension



Deliverable 1: the RDF conversion (section of RDF conversion)

```
<http://share-vde.org/sharevde/rdfBibframe2/Title/a0f4e860-c259-3611-b7ca-03b9369568cb> rdfs:label "Revelationes."  
<http://share-vde.org/sharevde/rdfBibframe2/uniform-title/a0f4e860-c259-3611-b7ca-03b9369568cb>  
  a ns1:Title ;  
  rdfs:label "Revelationes." ;  
  ns0:titleSortKey "Revelationes." ;  
  ns1:mainTitle "Revelationes." ;  
  ns0:title40MatchKey "Revelationes." ;  
  ns0:title40MarcKey "24010$aRevelationes. $lEnglish $s(Searby) $0http://id.loc.gov/authorities/names/n84007202" .
```

```
<http://share-vde.org/sharevde/rdfBibframe2/Instance/LOC13910411>  
  ns1:instanceOf <http://id.loc.gov/authorities/names/n84007202> ;  
  ns1:issuance <http://id.loc.gov/vocabulary/issuance/mono> ;  
  ns1:adminMetadata <http://share-vde.org/sharevde/rdfBibframe2/AdminMetadata/ad1517aa-fe91-3d10-902c-a0de7cbd787e>,  
<http://share-vde.org/sharevde/rdfBibframe2/AdminMetadata/4abffffde-8e4b-3806-b70a-e5c95af4f221>, <http://share-  
vde.org/sharevde/rdfBibframe2/AdminMetadata/11d5ffba-c1cd-3077-a2f5-aeb0b1fdc4f9> ;  
  ns1:contribution <http://share-vde.org/sharevde/rdfBibframe2/Contribution/8cd47f23-58b9-3a8c-b1f9-4627fca86c1a>,  
<http://share-vde.org/sharevde/rdfBibframe2/Contribution/e3d5c3a8-560a-3737-a105-dc7a4e8eadba> ;  
  a ns1:Instance ;  
  ns1:media <http://rdaregistry.info/termList/RDAMediaType/1007>, <http://id.loc.gov/vocabulary/mediaTypes/n> ;  
  ns1:carrier <http://rdaregistry.info/termList/RDACarrierType/1049>, <http://id.loc.gov/vocabulary/carriers/nc> ;  
  ns1:title <http://share-vde.org/sharevde/rdfBibframe2/title-statement/1835091e-3641-34e6-b0c5-fe64446acc12> ;  
  ns1:dimensions "25 cm" ;  
  ns1:extent <http://share-vde.org/sharevde/rdfBibframe2/Extent/0d4d8bca-a46a-3701-874b-8c45a15205c7> ;  
  ns1:note <http://share-vde.org/sharevde/rdfBibframe2/Note/2165960f-7bd9-3e33-b926-23c0de8c9765>, <http://share-  
vde.org/sharevde/rdfBibframe2/Note/92ace715-99d1-30a5-bca0-aec0d55a14ec> ;  
  ns1:responsibilityStatement "translated by Dennis Searby with Introduction and Notes By Bridget Morris." ;  
  ns1:provisionActivity <http://share-vde.org/sharevde/rdfBibframe2/ProvisionActivity/0c8dec7d-3f14-30cb-8074-  
7049da6859f9> ;
```



Deliverable 2: the Knowledge base of clusters - an example of clusterization algorithms

Clusterization of "forename" heading type

Example:

"\$aBridget,\$cof Sweden, Saint,\$dapproximately 1303-1373"

1) selections of interesting subfield

2) normalization of string text without diacritics, accents: Bridget of Sweden Saint approximately 1303-1373.

3) translate all in uppercase and search string into db variant forms and cross references:
BRIDGET OF SWEDEN SAINT APPROXIMATELY 1303-1373

4) if no cluster found, subfields will be analyzed

4.1 comparing \$a with other existing forms

4.2 comparing only the numeric part of \$d (having same \$a): \$dapproximately 1303-1373
=> \$d1303-1373

4.3 comparing \$c for "saint" or "santa" or other forms (having same \$a)

Deliverable 2: the Knowledge base of clusters - Postgres as a bridge

<input type="checkbox"/>	dstr_id integer	hdg_id text	name text	pref_fm boolean
<input type="checkbox"/>	151177	AN163420	Bridget, of Sweden, Saint, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX1253673	Bridget, of Sweden, Saint, approximately 1303-1373. ...	false
<input type="checkbox"/>	151177	AX1253674	Bridget, of Sweden, Saint, approximately 1303-1373. ...	false
<input type="checkbox"/>	151177	AX1253675	Bridget, of Sweden, Saint, approximately 1303-1373. ...	false
<input type="checkbox"/>	151177	AX1253676	Bridget, of Sweden, Saint, ca. 1303-1373. Uppenbarel...	false
<input type="checkbox"/>	151177	AX1253677	Bridget, of Sweden, Saint, ca. 1303-1373. Reuelation...	false
<input type="checkbox"/>	151177	AX1326216	Bridget, of Sweden, Saint, approximately 1303-1373. ...	false
<input type="checkbox"/>	151177	AX1326217	Bridget, of Sweden, Saint, ca. 1303-1373. Revelationes	false
<input type="checkbox"/>	151177	AX186404	Birgit, of Sweden, Saint, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX186405	Birgitta Birgersdotter, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX186406	Birgitta, of Sweden, Saint, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX186407	Birgitta, Saint, of Sweden, d. 1373	false
<input type="checkbox"/>	151177	AX186408	Birgitta, von Schweden, Saint, approximately 1303-13...	false
<input type="checkbox"/>	151177	AX186409	Bridget, of Sweden, Saint, ca. 1303-1373	false
<input type="checkbox"/>	151177	AX186410	Brigid, of Sweden, Saint, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX186411	Brigida, di Svezia, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX186412	Brigida, of Sweden, Saint, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX186413	Brigitta, of Sweden, Saint, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX186414	Brigitte, de Suède, Saint, approximately 1303-1373	false
<input type="checkbox"/>	151177	AX186415	Brigitte, of Sweden, Saint, approximately 1303-1373	false

Deliverable 2: the Knowledge base of clusters - the final result

▼ This person in

-  isni
-  WIKIDATA
-  LIBRARY OF CONGRESS
-  WorldCat Identities
-  data.bnf.fr
-  YAF



Birgitta, helgon, 1303-1373

ID: 151177

Works

<http://share-vde.org/sharevde/rdfBibframe/Agent/151177> <<http://www.w3.org/2002/07/owl#sameAs>>
<http://isni.org/isni/0000000121012842> .
<http://share-vde.org/sharevde/rdfBibframe/Agent/151177> <<http://www.w3.org/2002/07/owl#sameAs>>
<http://www.wikidata.org/entity/Q204996> .
<http://share-vde.org/sharevde/rdfBibframe/Agent/151177> <<http://www.loc.gov/mads/rdf/v1#isIdentifiedByAuthority>>
<http://id.loc.gov/authorities/names/n50044402> .

<<http://share-vde.org/sharevde/rdfBibframe/Agent/151177>>

rdfs:label "Bridget, of Sweden, Saint, approximately 1303-1373. Puch der himlischen Offenbarung der heiligen Wittiben Birgitte von dem Königreich Sweden", "Birgitta Suecica 1303-1373", "Bridget, of Sweden, Saint, ca. 1303-1373. Revelaciones", "Brigida, of Sweden, Saint, approximately 1303-1373", "Birgitta Birgersdotter, approximately 1303-1373", "Birgitta, of Sweden, Saint, approximately 1303-1373. Liber celestis of St. Bridget of Sweden", "Bridget, of Sweden, Saint, approximately 1303-1373. Revelations", "Birgitta, Saint, of Sweden, d. 1373", "Bridget, of Sweden, Saint, approximately 1303-1373. Saint Bride and her book", "Brigida da Suécia, Santa, 1303-1373", "Birgitta, helgon, 1303-1373", "Bridget, of Sweden, Saint, ca. 1303-1373", "Bridget, of Sweden, Saint, approximately 1303-1373. Reuelationes celestes praelecte sponse Christi beate Birgitte vidue de regno Suecie", "Bridget, of Sweden, Saint, ca. 1303-1373. Selections", "Brigida, de Suecia, Santa, Ca. 1303-1373", "Bridget, of Sweden, Saint, approximately 1303-1373. Revelations of Saint Birgitta", "Bridget, of Sweden, Saint, approximately 1303-1373. Brigida di Svezia", "Bridget, of Sweden, Saint, approximately 1303-1373", "Birgitta, von Schweden, Saint, approximately 1303-1373", "Bridget, of Sweden, Saint, ca. 1303-1373. Sermo angelicus", "Birgitta, ca. 1303-1373", "Bridget, of Sweden, Saint, approximately 1303-1373. Book V of St Birgitta's Uppenbarelsen" ;

▼ Other name forms

-  Bridget, of Sweden, Saint, approximately 1303-1373
-  Brigida, de Suécia, santa, 1303-1373
-  Brigitta, s., vidua, c. 1303-137
-  Brigita, Zviedrijas, Svētā, ap 1303-1373
-  Bridget, of Sweden, Saint, approximately 1303-1373
-  Birgitta святая 1303-ок.1373
-  Brigitta, de Suède, sainte
-  Brigida 1303-1373 de Suécia, santa
-  Bridget, of Sweden, Saint, ca. 1303-1373
-  Bridget (Sw.; 1302-1373)
-  Sweden, Saint, ely 1303-1373
-  Saint, of Sweden, d.
-  Bridget, von Schweden, Saint, approximately 1303-1373
-  Bridget, of Sweden, Saint, approximately 1303-1373
-  Brigitte, de Suède, Saint, approximately 1303-1373
-  Bride, of Sweden, Saint, ca. 1303-1373

...(other forms)



URI table for external sources

SHARE-VDE

URI_table_for_external_sources_2017-12-22.xls

2017-12-22

CODE	SOURCE	INSTITUTION	LINK TO PROJECT PAGE	URI	URI EXAMPLE	MARC21 TAG	API/WS
NAF (Names)	Library of Congress Names	Library of Congress	http://id.loc.gov/authorities/names.html	http://id.loc.gov/authorities/names	http://id.loc.gov/authorities/names/n79021285	100/110/111 \$n	IS/M

SHARE-VDE								URI_table_for_external_sources_2017-12-22.xls							
SPARQL endpoint	RDF dump	LICENSE TERMS	URL SEARCH		NOTES	AUTHIFY (Yes for search URL specified)		SUBFIELD		\$0/\$1					
	Yes		http://id.loc.gov/search/?q=Shakespeare+William&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fauthorities%2Fnames&format=atom&start=1			http://id.loc.gov/authorities/names									
	Yes		http://id.loc.gov/search/?q=bible+old+testament&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fauthorities%2Fnames OR http://id.loc.gov/search/?q=niccolo+macchiavelli+arte+della+guerra&q=cs:http://id.loc.gov/authorities/names			http://id.loc.gov/authorities/names/n79021285									
	Yes		http://id.loc.gov/search/?q=poetry&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fauthorities%2Fsubjects&format=atom&start=1			http://id.loc.gov/authorities/names/n80106714 http://id.loc.gov/authorities/names/nr00019197									
	NO		http://id.loc.gov/search/?q=poetry&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fauthorities%2Fclassifications&format=atom&start=1			NO									
	NO		http://id.loc.gov/search/?q=viaf&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fvocabulary%2Fidentifiers&format=atom&start=1												
			http://id.loc.gov/search/?q=italy&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fvocabulary%2FgeographicAreas&format=atom&start=1		It can be added as a fix value "http://id.loc.gov/vocabulary/geographicAreas/" + \$. For example: http://id.loc.gov/vocabulary/geographicAreas/a-ko	NO									
			http://id.loc.gov/search/?q=creator&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fvocabulary%2Frelators&format=atom&start=1		Addition of URI in export process: creation of a table that searches the corresponding URI from the \$e field. The /detect service exist, it retrieves the relator code from the responsibility expressed in the 245 \$c tag.	/detect									
			http://id.loc.gov/search/?q=italian&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fvocabulary%2Flanguages&format=atom&start=1			NO									
			http://id.loc.gov/search/?q=italy&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fvocabulary%2Fcountries&format=atom&start=1			NO									
			http://isni.oclc.nl/sru/?query=pica.na%3D%22Shakespeare%2C%20William%22&operation=searchRetrieve&recordSchema=isni-b&maximumRecords=3&recordPacking=xml		URI implementation with the available API. Massive usage with authentication.	NO									
	Yes		http://viaf.org/viaf/search?query=local.personalNames+%22Shakespeare%20William%22&sortKeys=holdingscount&maximumRecords=2&httpAccept=application/xml			http://id.loc.gov/vocabulary/relators/aut http://isni.org/isni/0000000120981297 http://viaf.org/viaf/14356									



Deliverable 4: enriched MARC21

Record example: <http://id.loc.gov/tools/bibframe/compare-id/full-ttl?find=13910411>

Original record (by LOC)

```
01996cam a2200421 i 4500
001 13910411
005 20161026123523.0
008 050324s2015 nyua b 001 0 eng
010 $a 2005047277
020 $a9780195166446 (v.1)
020 $a9780195166262 (v.2)
020 $a9780195166279 (v.3)
020 $a9780195166286 (v.4)
040 $aDLC$beng$cDLC$erda
041 1 $aeng$hlat
042 $apcc
050 00 $aBX4700.B62$bE5 2006
100 0 $aBridget,$cof Sweden, Saint,$dapproximately 1303-1373.
240 10 $aRevelationes.$lEnglish$s (Searby)
245 14 $aThe revelations of St. Birgitta of Sweden Volume 4 : the heavenly emperor's book to kings, the
rule, and minor works /$ctranslated by Dennis Searby with Introduction and Notes By Bridget Morris.
264 1 $aOxford :$bOxford University Press,$c2006-[2015]
300 $a4 volumes :$billustrations ;$c25 cm
```



Deliverable 4: enriched MARC21

Record example: <http://id.loc.gov/tools/bibframe/compare-id/full-ttl?find=13910411>

Original record (by LOC)

336 \$a\$text\$btxt\$2rdacontent
337 \$aunmediated\$bn\$2rdamedia
338 \$avolume\$bnc\$2rdacarrier
504 \$aIncludes bibliographical references and index.
505 0 \$a v. 1. Liber Caelestis, books I-III -- v. 2. Liber Caelestis, books IV-V -- v. 3. Liber Caelestis, books VI-VII -- v. 4. The heavenly emperor's book to kings, the rule, and minor works.
650 0 \$aPrivate revelations.
650 0 \$aVisions.
700 1 \$aSearby, Denis Michael,\$etranslator.
700 1 \$aMorris, Bridget,\$d1954-\$ewriter of supplementary textual content.
906 \$a0\$bibc\$corignew\$d2\$encip\$f20\$gy-gencatlg
925 0 \$aacquire\$b1 shelf copy\$xpolicy default
952 \$aComplete in 4 vols.
955 \$brm08 2016-06-06 (Telework)
955 \$apc16 2005-03-24 to HLCD\$csch21 2005-03-30\$dsh13 2005-03-31\$esh42 2005-04-04 to Dewey\$aaa05 2005-04-06\$aps12 2006-07-19 1 copy rec'd., to CIP ver.\$fpv17 2006-08-03 Z-CipVer\$arc09 2009-06-25 v. 2 added\$trf13 2009-07-27 c. 2, to BCCD
955 \$aADDED VOLS: v. 3 xn05 2012-4-5 to USGEN
955 \$aADDED VOLS: v. 4 xn12 2015-12-01 to USASH



Deliverable 4: enriched MARC21

Record example: <http://id.loc.gov/tools/bibframe/compare-id/full-ttl?find=13910411>

Record enriched with URIs (LOC)

```
=LDR 02447cam a2200421 i 4500
=001 13910411
=005 20161026123523.0
=008 050324s2015\\nyua\\b\\001\0\eng\\
=010 \\$a 2005047277
=020 \\$a9780195166446 (v.1)
=020 \\$a9780195166262 (v.2)
=020 \\$a9780195166279 (v.3)
=020 \\$a9780195166286 (v.4)
=040 \\$aDLC$beng$cDLC$erda
=041 1$aeng$hlat
=042 \\$apcc
=050 00$aBX4700.B62$bE5 2006
=100 0$aBridget,$cof Sweden, Saint,$dapproximately 1303-1373.
$0http://id.loc.gov/authorities/names/n50044402
=240 10$aRevelationes.$lEnglish$s (Searby) $0http://id.loc.gov/authorities/names/n84007202
=245 14$aThe revelations of St. Birgitta of Sweden Volume 4 : the heavenly emperor's book to kings, the
rule, and minor works /$ctranslated by Dennis Searby with Introduction and Notes By Bridget Morris.
=264 \1$aOxford :$bOxford University Press,$c2006-[2015]
=300 \\$a4 volumes :$billustrations ;$c25 cm
```



Deliverable 4: enriched MARC21

Record example: <http://id.loc.gov/tools/bibframe/compare-id/full-ttl?find=13910411>

Record enriched with URIs (LOC)

```
=336  \\\$atext$btxt$2rdacontent$0http://rdaregistry.info/termList/RDAContentType/1020
=337  \\\$aunmediated$bn$2rdamedia$0http://rdaregistry.info/termList/RDAMediaType/1007
=338  \\\$avolume$bn$2rdacarrier$0http://rdaregistry.info/termList/RDACarrierType/1049
=504  \\\$aIncludes bibliographical references and index.
=505  0\$\av. 1. Liber Caelestis, books I-III -- v. 2. Liber Caelestis, books IV-V -- v. 3.
Liber Caelestis, books VI-VII -- v. 4. The heavenly emperor's book to kings, the rule, and minor works.
=650  \0$aPrivate revelations.$0http://id.loc.gov/authorities/subjects/sh85107042
=650  \0$aVisions.$0http://id.loc.gov/authorities/subjects/sh85143882
=700  1\$\aSearby, Denis Michael,$etranslator.$0http://id.loc.gov/authorities/names/nr98021028
=700  1\$\aMorris, Bridget,$d1954-$ewriter of supplementary textual
content.$0http://id.loc.gov/authorities/names/n92016617
=906  \\\$a0$bibc$corignew$d2$encip$f20$gy-gencatlg
=925  0\$\aacquire$b1 shelf copy$xpolicy default
=952  \\\$aComplete in 4 vols.
=955  \\\$brm08 2016-06-06 (Telework)
=955  \\\$apc16 2005-03-24 to HLCD$csh21 2005-03-30$dsh13 2005-03-31$esh42 2005-04-04 to Dewey$aaa05 2005-04-06$aps1
2006-07-19 1 copy rec'd., to CIP ver.$fpv17 2006-08-03 Z-CipVer$arc09 2009-06-25 v. 2
added$trf13 2009-07-27 c. 2, v. 2 to BCCD
=955  \\\$aADDED VOLS: v. 3 xn05 2012-4-5 to USGEN
=955  \\\$aADDED VOLS: v. 4 xn12 2015-12-01 to USASH
```



SHARE-VDE

possible steps for the future



Candidate Use Cases for a production phase

Phase 3a

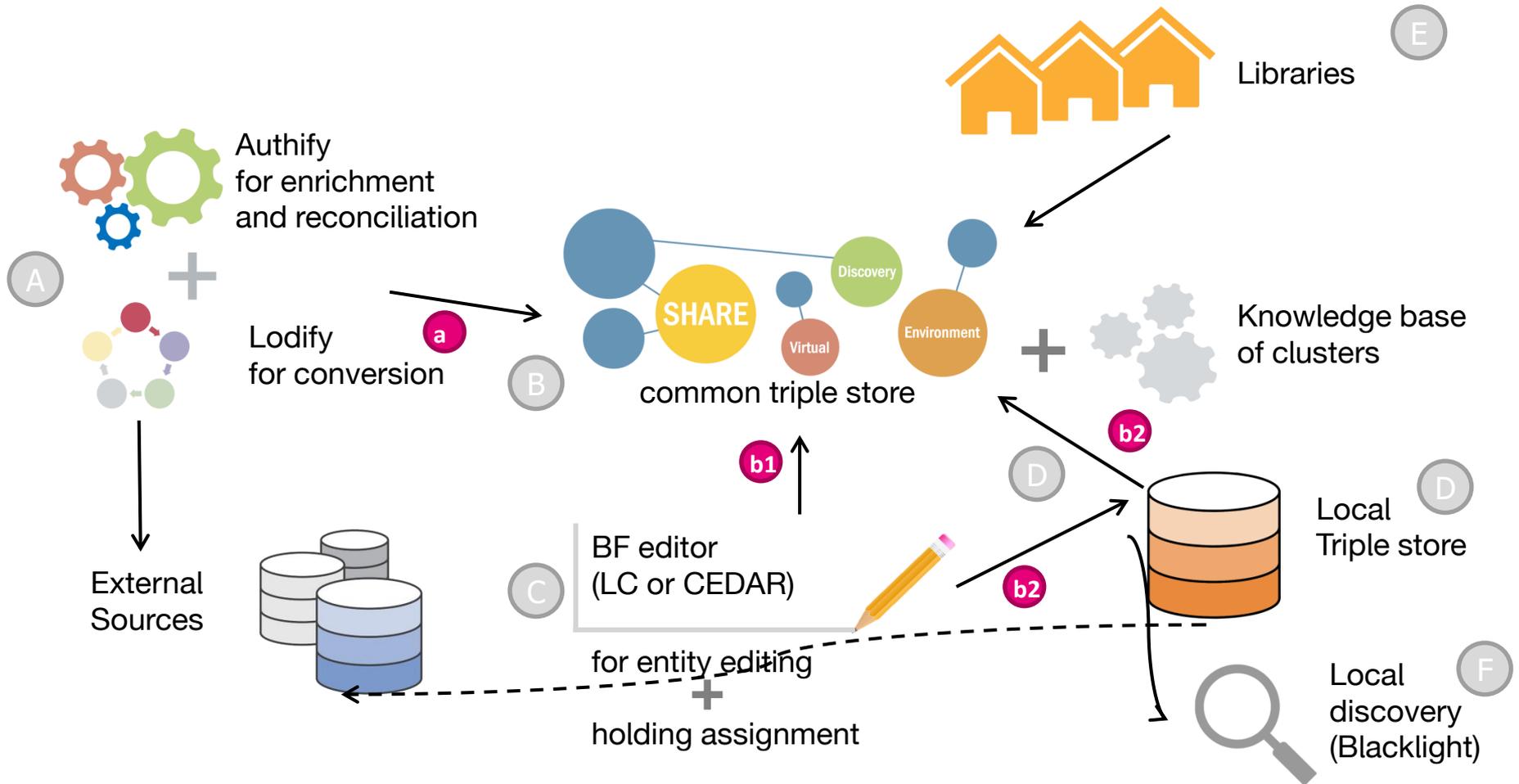
- Publication of the entire catalogues on the SHARE-VDE platform, updated SHARE-VDE common knowledge base [UC1]
- Batch or automated updating of data from libraries [UC2]
- Dissemination of data to contributing libraries on automated or batch process [UC3]

Phase 3b

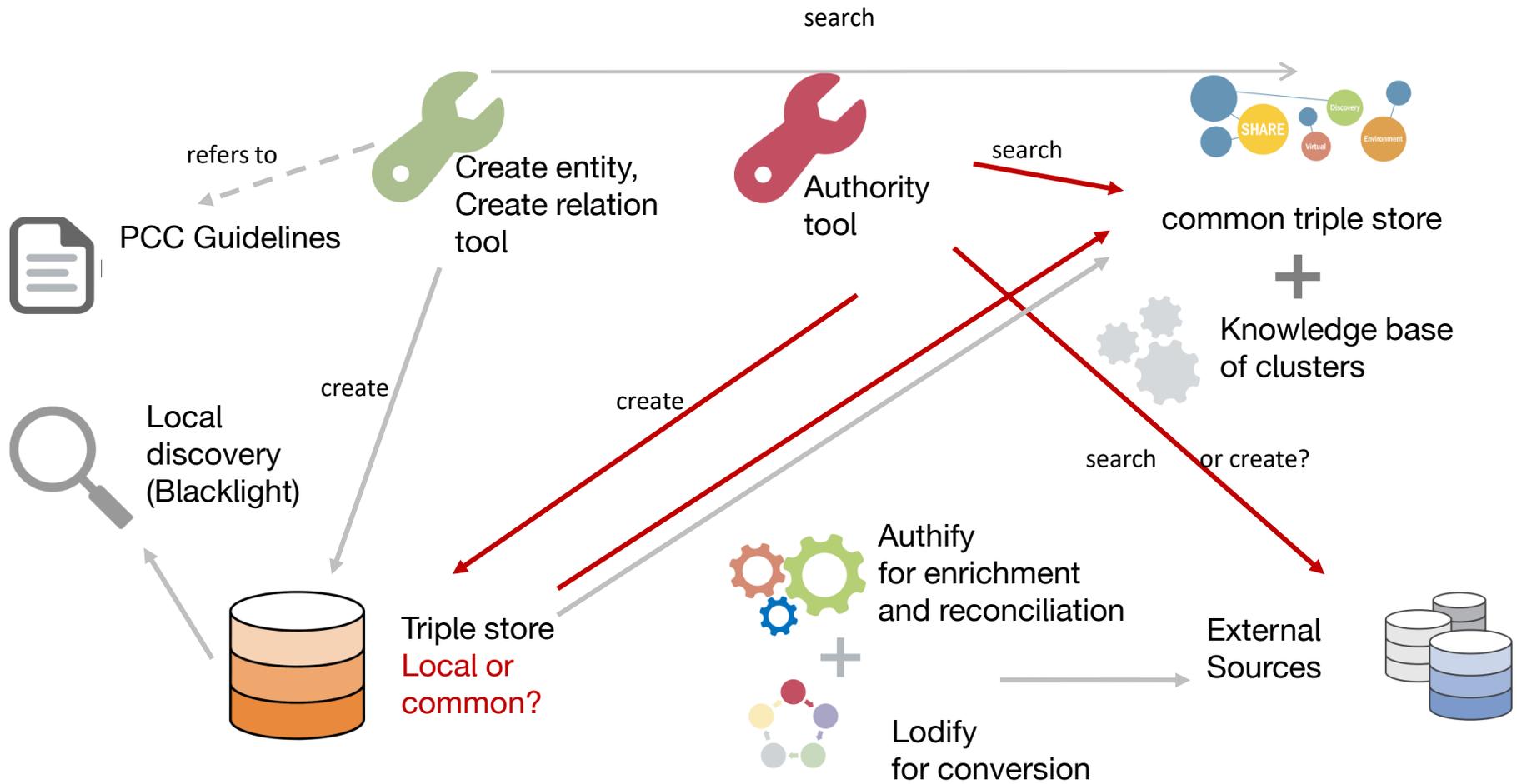
- Interaction with the common knowledge base [UC4]
- Reporting to serve library needs [UC5]
- Engage in cataloguing activities (holding assignment, entity editing, entity creation) using third party cataloguing tools [UC6]



Copy cataloguing



Original cataloguing



● Linkage to external authorities and web context data



Conclusions: the sharing and reuse of information resources

All of these efforts are being made with the aim of facilitating the **sharing and reuse of assets, and tools** produced by libraries, museums and other institutions, guaranteeing their availability to a wider public, enriching the World Wide Web with valuable information that would otherwise remain mostly hidden in archives, collections and catalogues, and promoting a **culture of open access to knowledge**, with numerous advantages for each link in the information chain.

Libraries, archives and museums all benefit from the possibility of more comprehensive and well-structured tools which provide end users with a **vast wealth of information**, and **create new co-operative tools** for information professionals.

In line with this new, open philosophy of data sharing and reuse, **even traditional authority controls, union catalogues and discovery systems are evolving.**



The Library of Congress
LC's Digital Future and You!
a series of briefings sponsored by Library Services on digital initiatives

Tuesday, February 6, 2018

Thank you!

Any questions or feedback are greatly appreciated.

Michele Casalini
Casalini Libri

Tiziana Possemato
Casalini Libri - @Cult

