Subject Cataloguing – Quo vadis?
Satellite conference associated with the annual meeting of the RDA Steering Committee RSC
4 November 2016 - German National Library in Frankfurt/Main

The central role of URIs in Subject Cataloguing

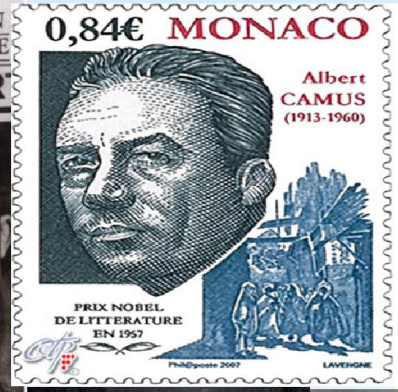
Tiziana Possemato
Casalini Libri | @Cult

Who's Who?

The problem to be solved:

how to identify an Entity?

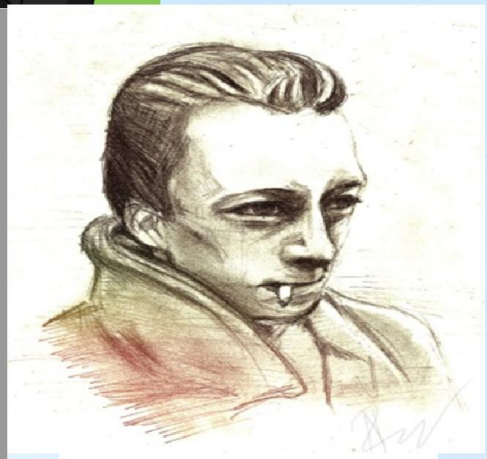

COMBAT
LE JOURNAL DE PARIS
ÉDITION SPÉCIALE
ALBERT CAMUS EST MOR
Une conscience contre le chaos

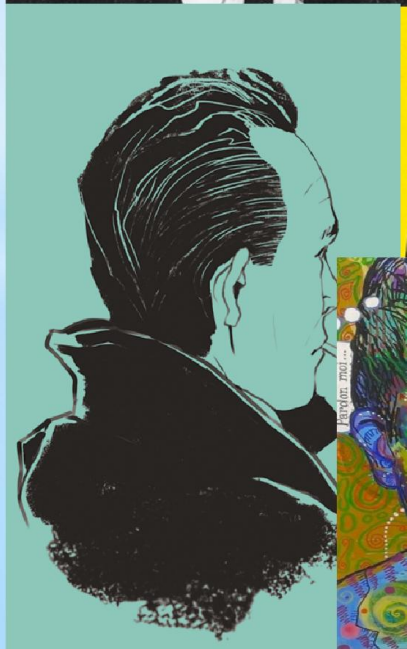

0,84€ MONACO
Albert
CAMUS
(1913-1960)
PRIX NOBEL
DE LITTERATURE
EN 1957

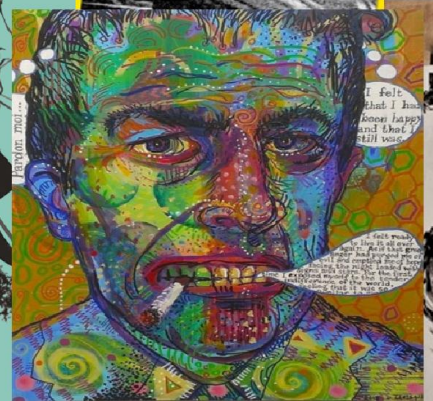

BACK TO YOU, BEAUVOIR.


BIBLIOTHEQUE
A.R.T.


I don't care, and you
find that attractive.
But I also don't care about that.
Happy Valentine's Day ♡
Love, _____ and
Albert Camus
benkling.tumblr







Albert Camus

$0 http://share_cat/person/131426

Search Person/Family/Corporate body

EXPAND ALL CLOSE ALL

This person in

isni
WIKIDATA
LIBRARY OF CONGRESS
data.bnf.fr
VIAF

Wikipedia

Albert Camus (French: [albɛʁ kamy]; 7 November 1913 – 4 January 1960) was a French philosopher, author, and journalist. His views contributed to the rise of the philosophy known as absurdism. He wrote in his essay *The Rebel* that his whole life was devoted to opposing the philosophy of nihilism while still delving deeply into individual freedom. He won the Nobel

Camus, Albert, 1913-1960
ID: 131426

Works

Other name forms

- Camus, Albert <1913-1960>
- كامو، البير
- Albert Camus écrivain et journaliste français
- 1913-1960 ,קמי, אלבר
- Camus, Albert
- Camus, Albert, 1913-1960
- Камю, А. 1913-1960 Альбер
- Camus, A. 1913-1960 Albert
- Камю, Альбер, 1913-1960
- 1960-1913 ،كامو، ألبير
- Camus, Albert
- Camus, Albert <1913-1960>
- CAMUS, Albert
- CAMUS, Albert <1913-1960>

SHARE Catalogue project

The **SHARE Catalogue** project involved conversion to linked data, publication on different sites and open data portals and the building of a common portal to search data:

Libraries involved in the project:

- Università degli Studi di Napoli Federico II (Naples)
- Università degli Studi di Napoli L'Orientale (Naples)
- Università degli Studi Napoli Parthenope (Naples)
- Università degli Studi di Salerno (Salerno)
- Università degli Studi del Sannio (Benevento)
- Università degli Studi della Basilicata (Potenza)
- Università degli Studi del Salento (Lecce)

General description of the catalogues in SHARE Catalogue project

The original data come from three different LMS (Aleph, Sebina and Millennium), in Unimarc and Marc 21 formats.

Records used in the project:

- bibliographic records
- authority files

SHARE Catalogue project

Project aim: to integrate the considerable knowledge base represented by the universities' authority and bibliographic catalogues to enrich it with the new and in-flux one generated by the web, creating an integrated information system to provide users with a single access tool for the various Libraries' OPAC.

SHARE Catalogue

Project objectives:

- the conversion of data in RDF
- the enrichment of data through a connection to external projects, above all related to Authority files and online encyclopedias (e.g. VIAF, Library of Congress Name Authority file, Library of Congress Subject Headings, ISNI, Fast, Wikipedia, Wikidata, …)
- the publication of data in a three layer portal, following the BIBFRAME model (Work/Person – Instances – Items)
- *(in the next future)* the addition of Subjects to enrich Entities

BIBFRAME – Bibliographic Framework Initiative

The **Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services** document published by the Library of Congress on November 21, 2012, sets out a new data model designed as an evolution, in linked open data, of the Marc 21 format.

The reflections on the new cataloguing rules focus on some specific points, including:

- a greater level of identification and analysis of the data;
- greater attention to controlled vocabularies;
- more widespread use of terms instead of codes;
- emphasis on relationships;
- greater flexibility in controlled items.

BIBFRAME – Data model v. 2.0

"In translating the MARC 21 format to a Linked Data model it is important to deconstruct and then reconstruct the informational assets that comprise MARC". The BIBFRAME Model, version 2.0 (published on 2016, 21th of April) consists of the following core classes:

Work: The highest level of abstraction, a Work, in the BIBFRAME context, reflects the conceptual essence of the cataloged resource: authors, languages, and what it is about (subjects).

Instance: A Work may have one or more individual, material embodiments, for example, a particular published form. These are Instances of the Work. An Instance reflects information such as its publisher, place and date of publication, and format.

Item: An item is an actual copy (physical or electronic) of an Instance. It reflects information such as its location (physical or virtual), shelf mark, and barcode.

BIBFRAME – Data model v. 2.0

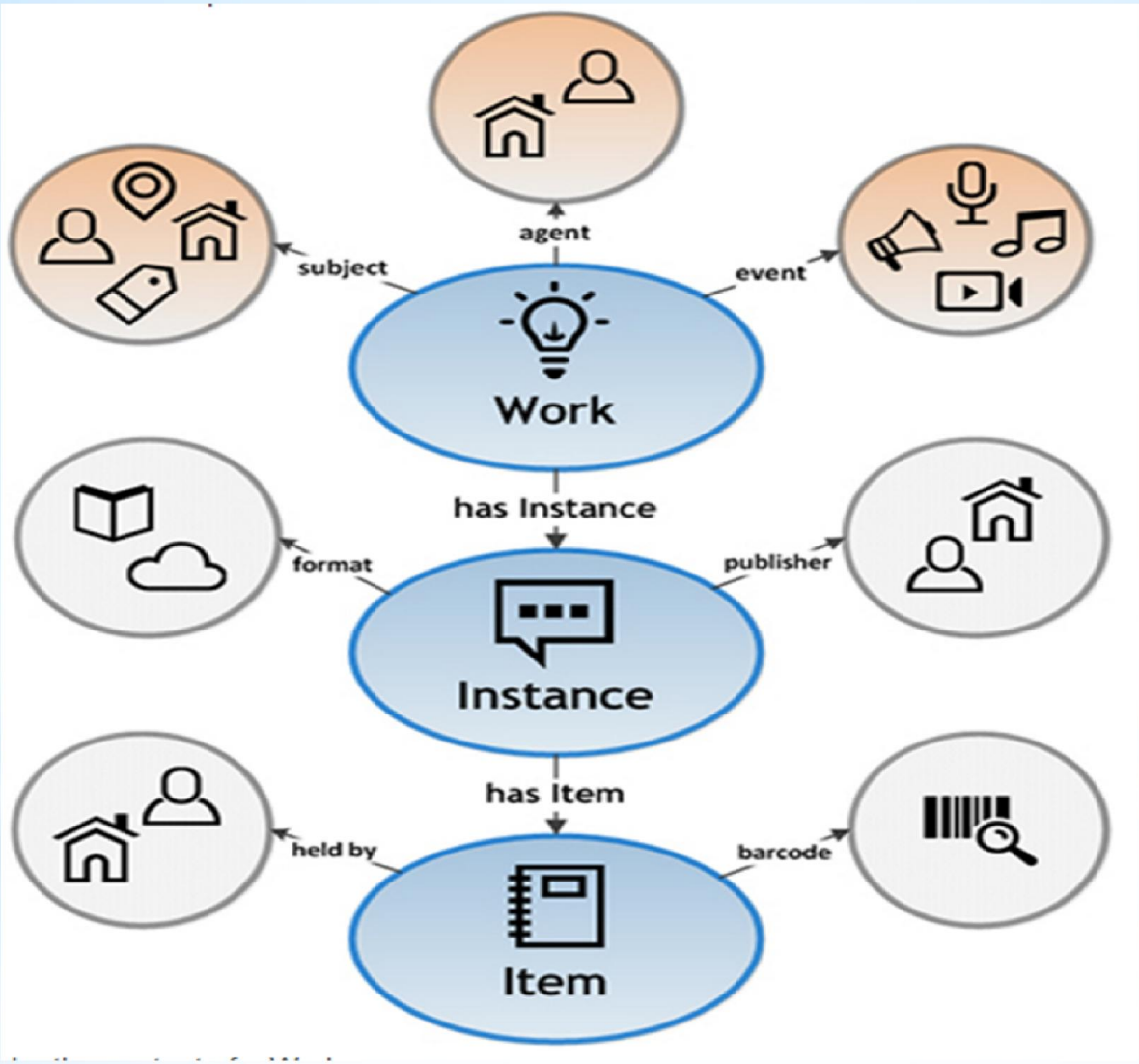
BIBFRAME 2.0 further defines additional key concepts that have relationships to the core classes:

Agents: Agents are people, organizations, jurisdictions, etc., associated with a Work or Instance through roles such as author, editor, artist, photographer, composer, illustrator, etc.

Subjects: A Work might be "about" one or more concepts. Such a concept is said to be a "subject" of the Work. Concepts that may be subjects include topics, places, temporal expressions, events, works, instances, items, agents, etc.

Events: Occurrences, the recording of which may be the content of a Work

BIBFRAME – Data model v. 2.0



General description of the catalogues in SHARE project

Survey results:

- little or no widespread use of attributes useful for the identification of the Work and Expression entities (uniform titles, original titles and so on);
- little or no widespread use of authorized access points for Subjects, useful to identify Entity

General description of the catalogues

This first analysis of data suggested another approach:

- a study of the **Person-Family-Body** entities with their attributes;
- a comparison of the same Person* in different catalogues
- a study of the **Works** related to a Person

to study how to reconciliate these entities (Person and Work, at the moment)

(* From this point onwards, by *Person*, we refer to an entity as Person/Family/Body)

Entity Enrichment from external sources

For each Person entity, all the forms through which it can be represented are recovered from local authority files and from external authority sources (such as VIAF).

The enrichment of the original data takes place in different steps:

- Use of authority data (authorized and variant forms)
- Use of bibliographic data (data reported as creator in the record)
- Comparison with variant forms of name with external sources
- Creation of the final cluster, with a unique local identifier.

Entity Reconciliation

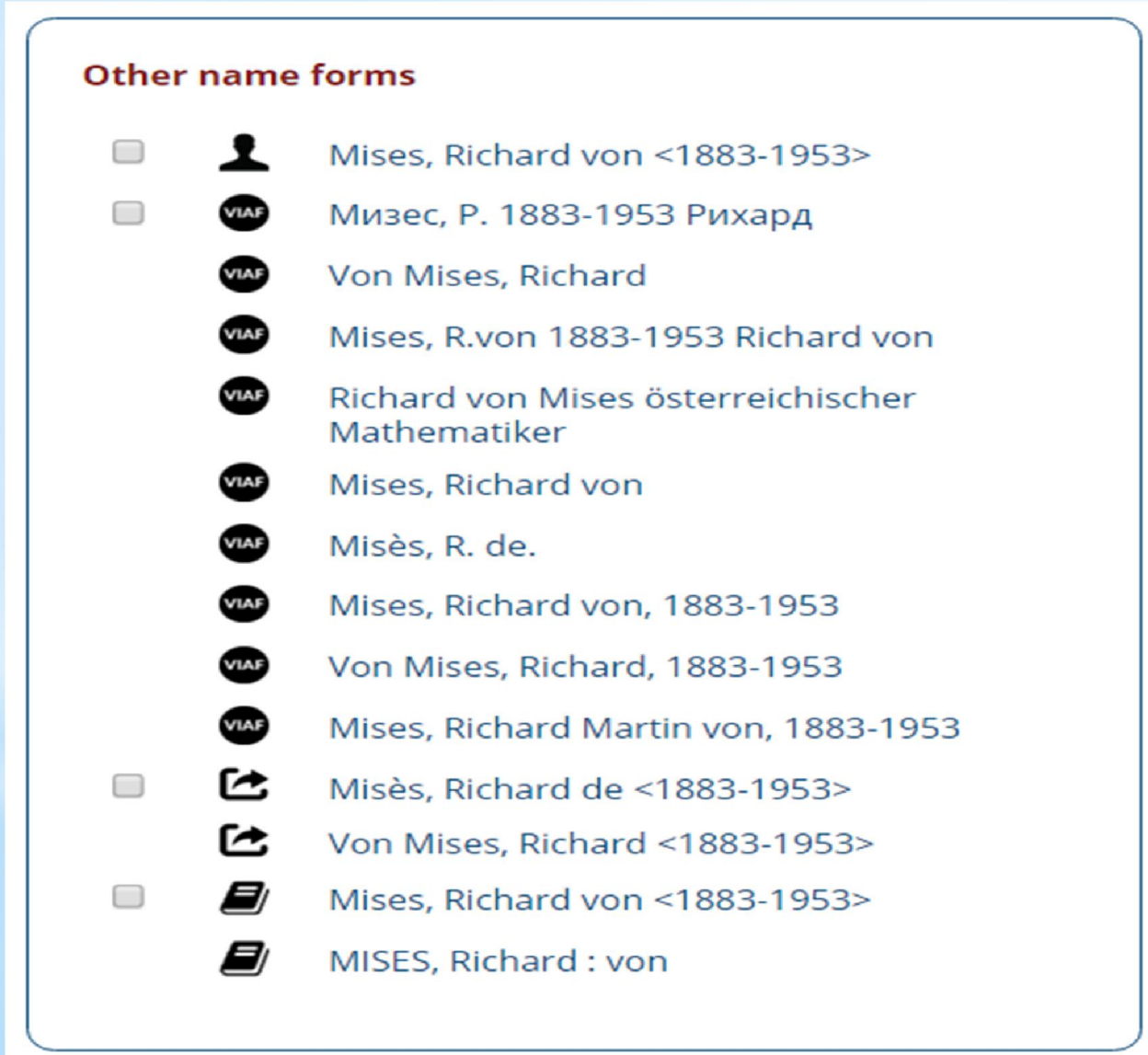
To obtain, as result:

- **cluster of names** – the forms, authorised and variants, of names of creator;
- **cluster of titles** – authorised access point and variant forms for the titles of the Works.

As second step:

- **cluster of subjects** - variant forms of a subject in different languages and other equivalent forms

Person cluster



Results of a Person cluster for the *Richard von Mises* entity with the different name forms stemming from:

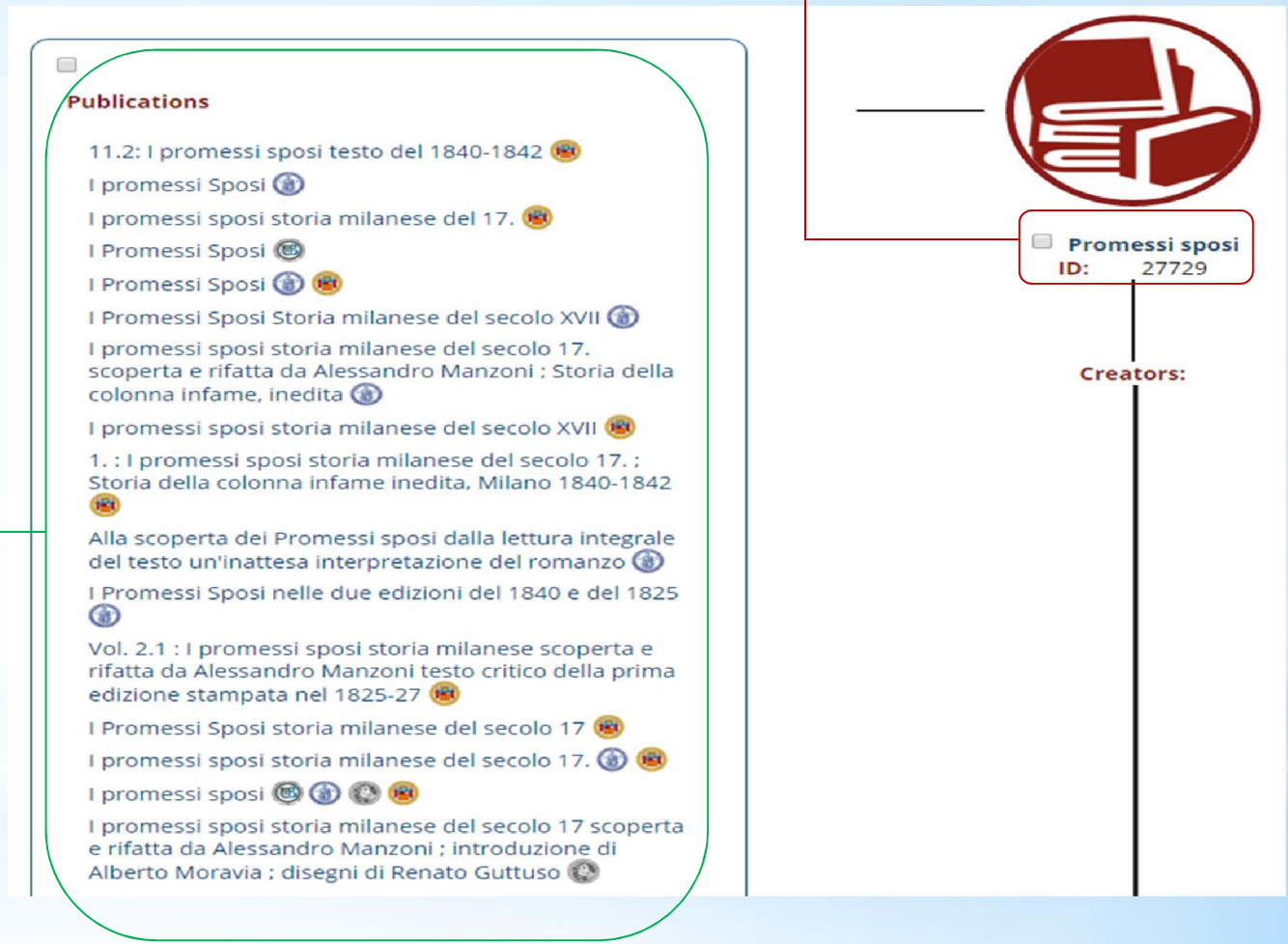
- Authority files
- External sources (s.a. VIAF)
- Referrals (to authority record)
- Forms used in bibliographic records

Title cluster

Grouping under a single work title of the many publication titles in the catalogue for *Promessi sposi*.

One work title

Brings together more than 70 different publications catalogued by the different libraries, which the end user can access with just one search.



Data reconciliation

The power of reconciliation:

- Search for Shakespeare *
- Search for Anna Karenina *

A Person cluster (William Shakespeare)

This person in

WIKIDATA

LIBRARY OF CONGRESS

data.bnf.fr

VIAF

Wikipedia

William Shakespeare (/ˈʃeɪkspɪər/; 26 April 1564 (baptised) – 23 April 1616) was an English poet, playwright, and actor, widely regarded as the greatest writer in the English language and the world's pre-eminent dramatist. He is often called England's national poet, and the "Bard of Avon". His extant works, including collaborations, consist of approximately 38 plays, 154 sonnets, two long narrative poems, and a few other verses, some of

Shakespeare, William, 1564-1616.
ID: 132200

Works

Other name forms

- Shakespeare, William <1564-1616>
- Shakespeare, William
- 1616-1564 ،وليم ،شكسبير
- Шекспир, У. 1564-1616 Уильям
- Shakespeare, William (English playwright and poet, 1564-1616)
- William Shakespeare
- Шекспир, Уильям, 1564-1616
- Shakespeare, W. 1564-1616 William
- Shakespeare, William, 1564-1616.
- 1564-1616 ,וליאם ,שיקספיר
- Shakespeare, William
- شكسبير ، وليام
- Shakespeare, Guglielmo

A Work cluster (Anna Karenina)

Publications

- Anna Karenina
- Anna Karenina Roman v vos'mi castjach
- v. 3. : Anna Karenina ; Racconti coniugali
- Ana Karenina
- Anna Karenina roman v'vos'mi castjah
- Anna Karenina roman v vos'mi castjah
- 8. : Anna Karenina roman v vosmi castjah
- Anna Karenina Cernovye redakcii i varianty
- 9. : Anna Karenina roman v vosmi castjah
- Anna Karenina Roman' v vos'mi castjach
- Anna Karenina Roman v vos'mi cast'jach
- Anna Karenina romanzo
- Anna Karenina Taccuini per "Anna Karenina"

Anna Karenina
ID: 21962

Creators:

(159663) Tolstoj, Lev Nikolaevič 1828-1910

Other forms

- Tolstoy, Leo, graf, 1828-1910. | Anna Karenina
- Tolstoï, Lev Nikolaevitch (1828-1910). | Anna Karénine
- アンナ・カレーニナ
- Толстой, Лев. | Анна Каренина : роман
- Tolstoj, Lev Nikolaevič | Ana Karenjina
- Tolstoï, Léon, 1828-1910. | Anna Karenina
- טולסטוי, לב ניקולויץ', 1828-1910. | אנה קרנינה
- Tolstoj, Lev Nikolaevič 1828-1910 | Anna Karenina

The next step: add the Subject cluster

Search Person/Family/Corporate body

EXPAND ALL | CLOSE ALL

This person in

isni
WIKIDATA
LIBRARY OF CONGRESS
data.bnf.fr
VIAF

Wikipedia

Karl Theodor Jaspers (German: [ˈjaspɐs]; 23 February 1883 – 26 February 1969) was a German-Swiss psychiatrist and philosopher who had a strong influence on modern theology, psychiatry, and philosophy. After being trained in and practicing psychiatry, Jaspers turned to philosophical inquiry and attempted to discover an innovative philosophical system. He was often viewed as a major exponent of existentialism in Germany, though he did not accept this label.

Wikipedia

Biography and activity

Filosofo

Jaspers, Karl, 1883-1969
ID: 123629

Works

Other name forms

- Jaspers, Karl <1883-1969>
- Jaspers, Karl, 1883-1969
- Ясперс, К. 1883-1969 Карл
- Jaspers, Karl
- JASPERS, KARL Dr
- 1883-1969 ,יספרס, קרל
- Karl Jaspers deutscher Philosoph und Psychiater
- 1969-1883 ,یاسپرز، کارل
- Jaspers, Karl <1883-1969>
- JASPERS, Karl
- Jaspers, Karl

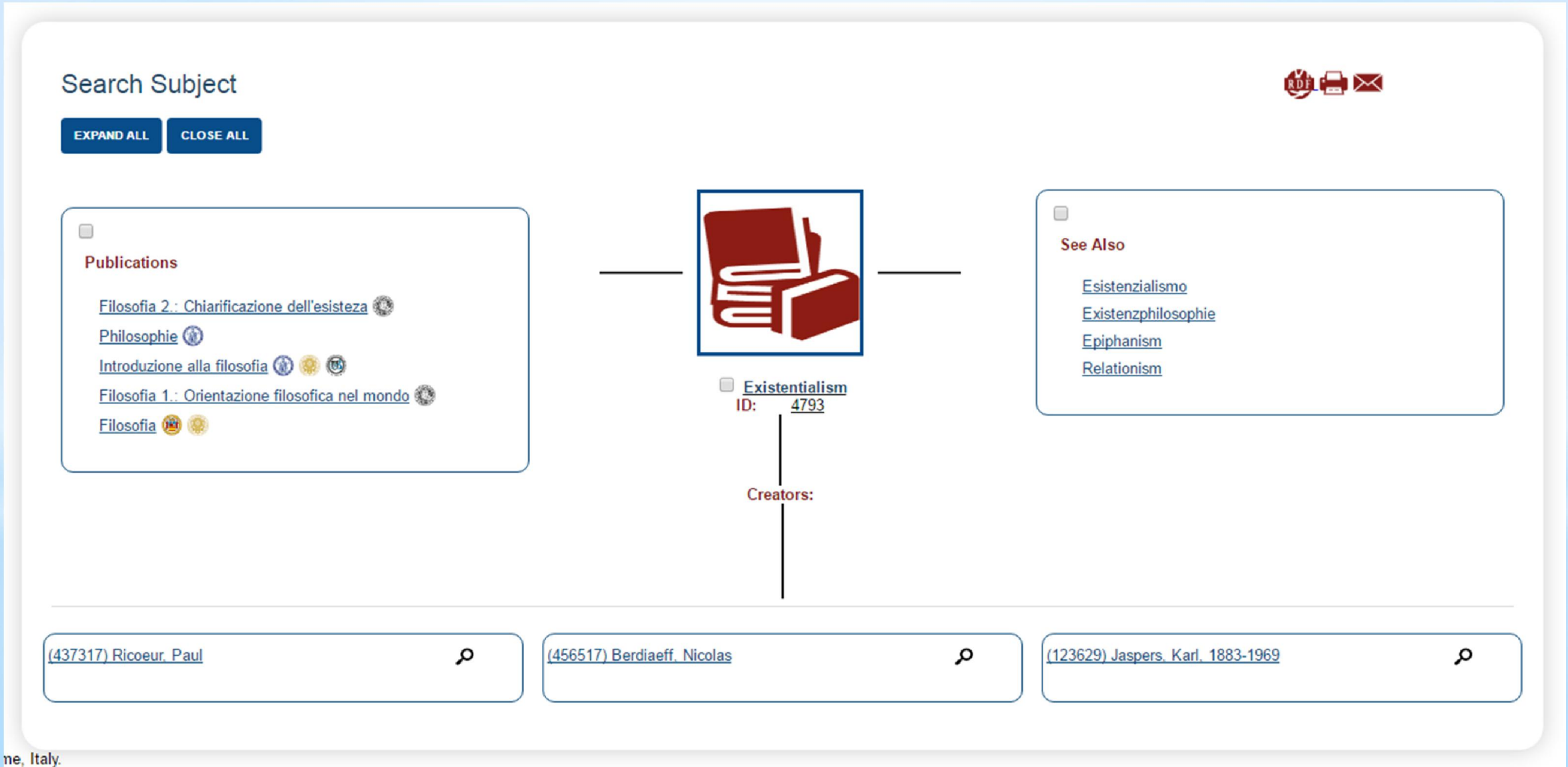
Subject

- Existentialism
- Philosophy
- Religion
- World War 1939 - 1945
- Germany - history

others...

See here the entity Karl Jasper enriched with subjects

The next step: add the Subject cluster



ne, Italy.

How we reconciliate entities

- Automated processes

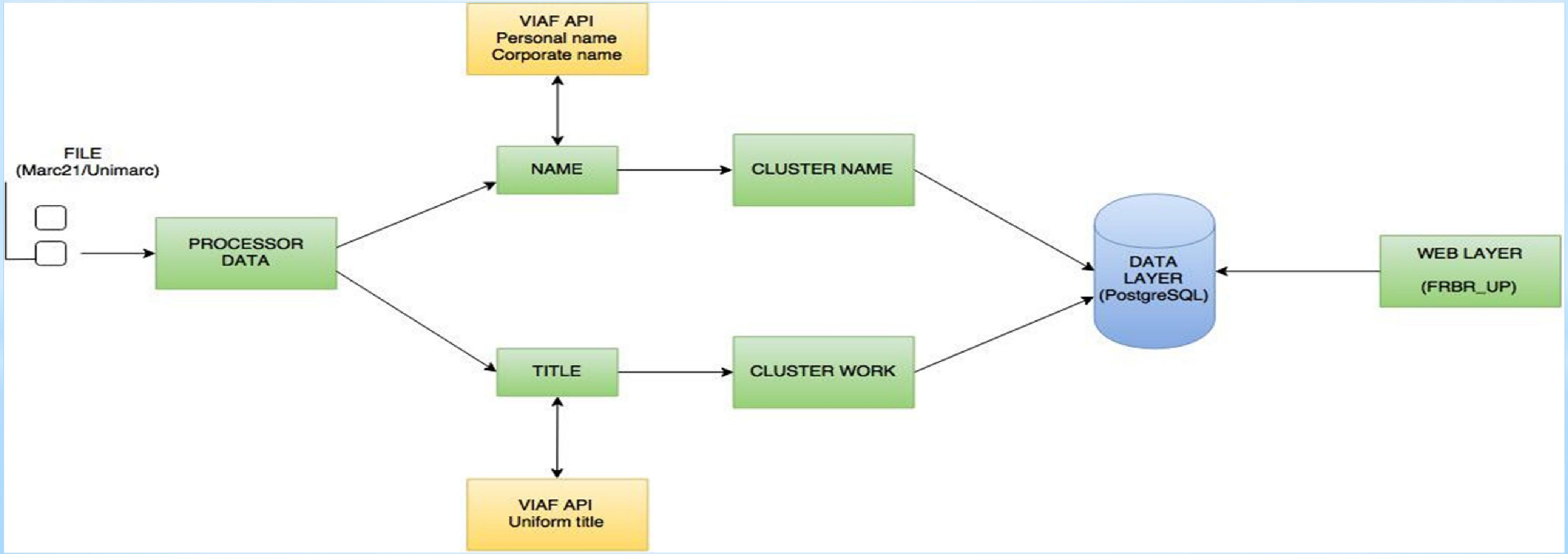
- Manual processes (in cataloguing flow)

Take into account how relations between reconciliation and validation may change in the automated or manual processes:

- automated processes: high level of clustering; low level of validation
- manual processes: low level of clustering; high level of validation

Automated processes to produce clusters

The loading processor and creation of Person/Work clusters: an important step of the process retrieves data from external authority files, such as VIAF, LCSH, ISNI, FAST using the specific APIs.



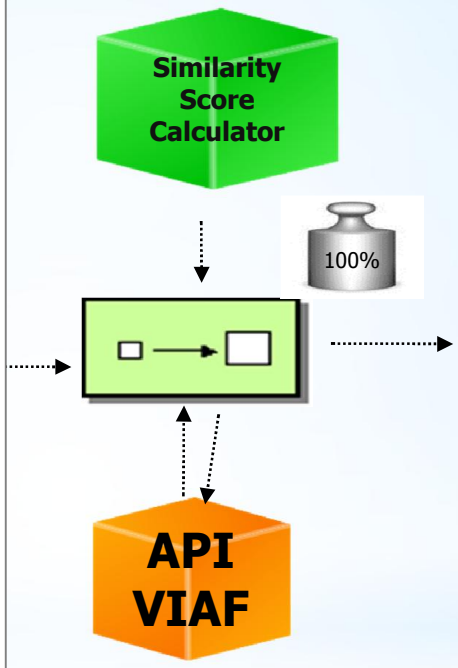
Cluster makers – Person (example 1)

```
001 00001
200 \1  $aEco,$bUmberto$f<1932- >.
400  \0 $aDedalus
997  \\ $aAUTHORITY
```

```
001 27283
700 \1  $aEco,$bUmberto$f<1932- >.
997 \\  $aUNINA
```

```
001 7258
700 \1  $aECO,$bUmberto
997 \\  $aUNISA
```

```
001 7258
700 \1  $aEco,$bUmberto
997 \\  $aUNIBAS
```



ID cluster: 1281829
Author : Eco, Umberto, 1932-...
VIAF ID: 108299403

Other forms:

ECO, Umberto
Eco, Umberto
Dedalus
1932- اكو، أمبيرتو
إكو، أمبرتو،
Eco, Umberto (Italian writer, architect, and critic, born 1932)
Umberto Eco semiologo, filosofo e scrittore italiano
1932 ,אקו, אומברטו-

This chart and the following example show the mechanism for associating names from different records in a single Person cluster

Cluster makers – Person (example 2)

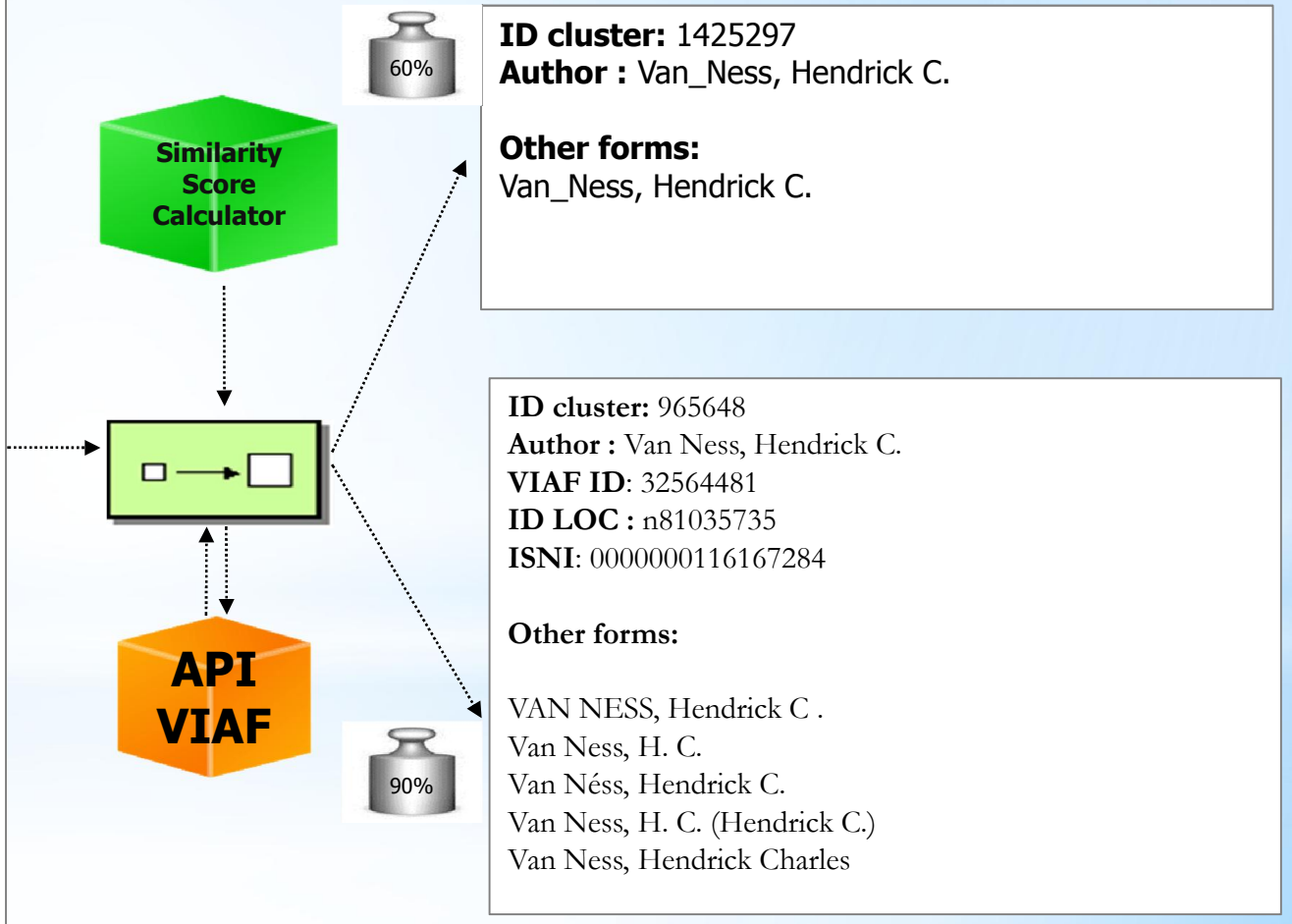
001 00002
200 \1 $aVan Ness,$bHendrick C.
997 \\ $aAUTHORITY

001 8379
701 \1 $aVan_Ness,$bHendrick C.
997 \\ $aUNIBAS

001 173506
701 \1 $aVan Ness,$bHendrick C.
997 \\ $aUNINA

001 1317
701 \1 $aVan Néss$b, Hendrick C.
997 \\ $aUNISANNIO

001 56522
701 \1 $aVAN NESS,$bHendrick C
997 \\ $aUNISA

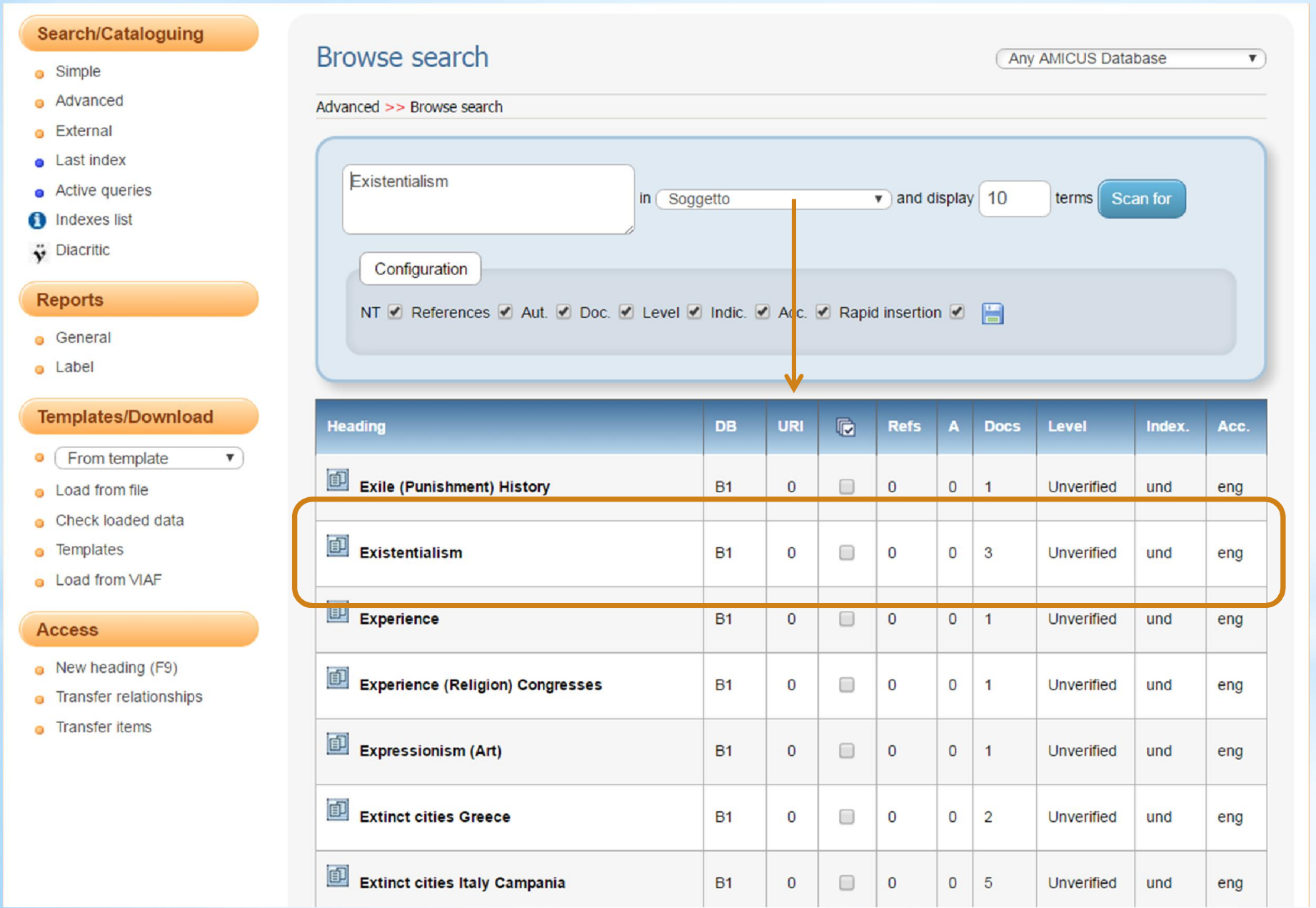


Manual process to produce entities clusters

We improved in the WeCat cataloguing module of OLISuite a «URI Management System», to manage identifiers for each access point or heading (names, titles, subjects, classifications etc.).

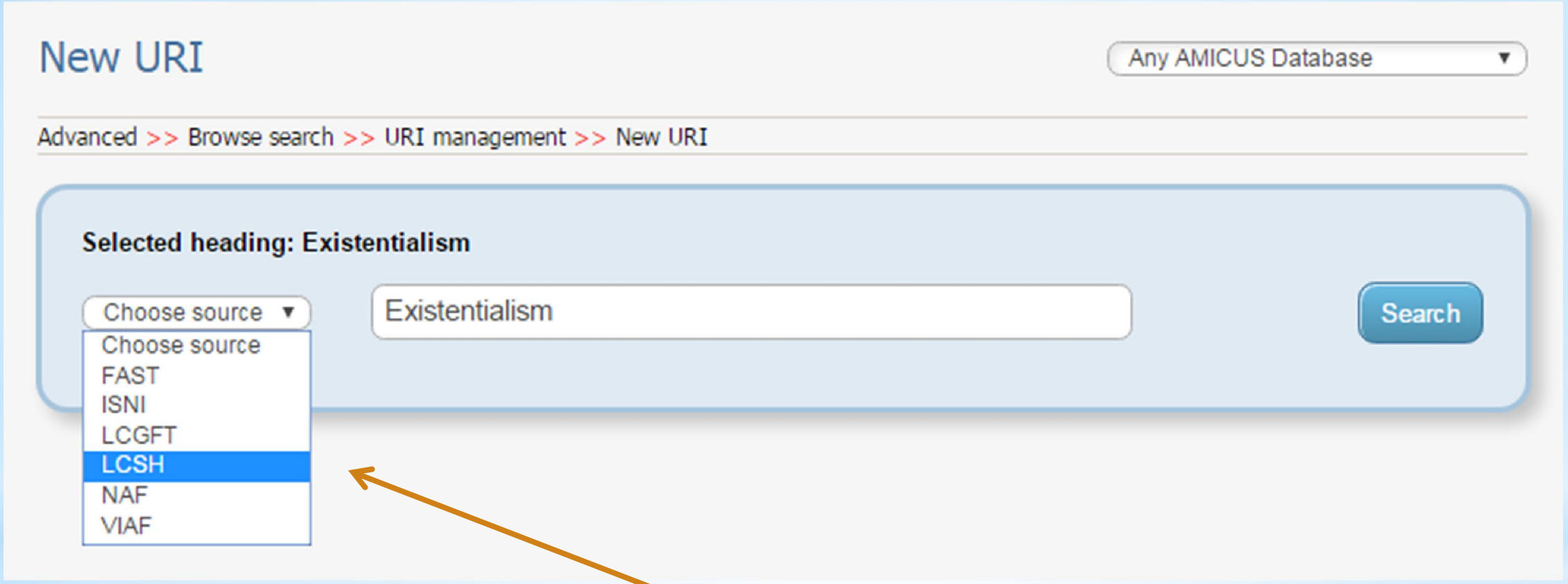
See as an example in the following slides: the authorized access point for *Existentialism* and, in the third column (URI), the number of URIs associated to the heading

URI Management System (OLIsuite/WeCat screen)



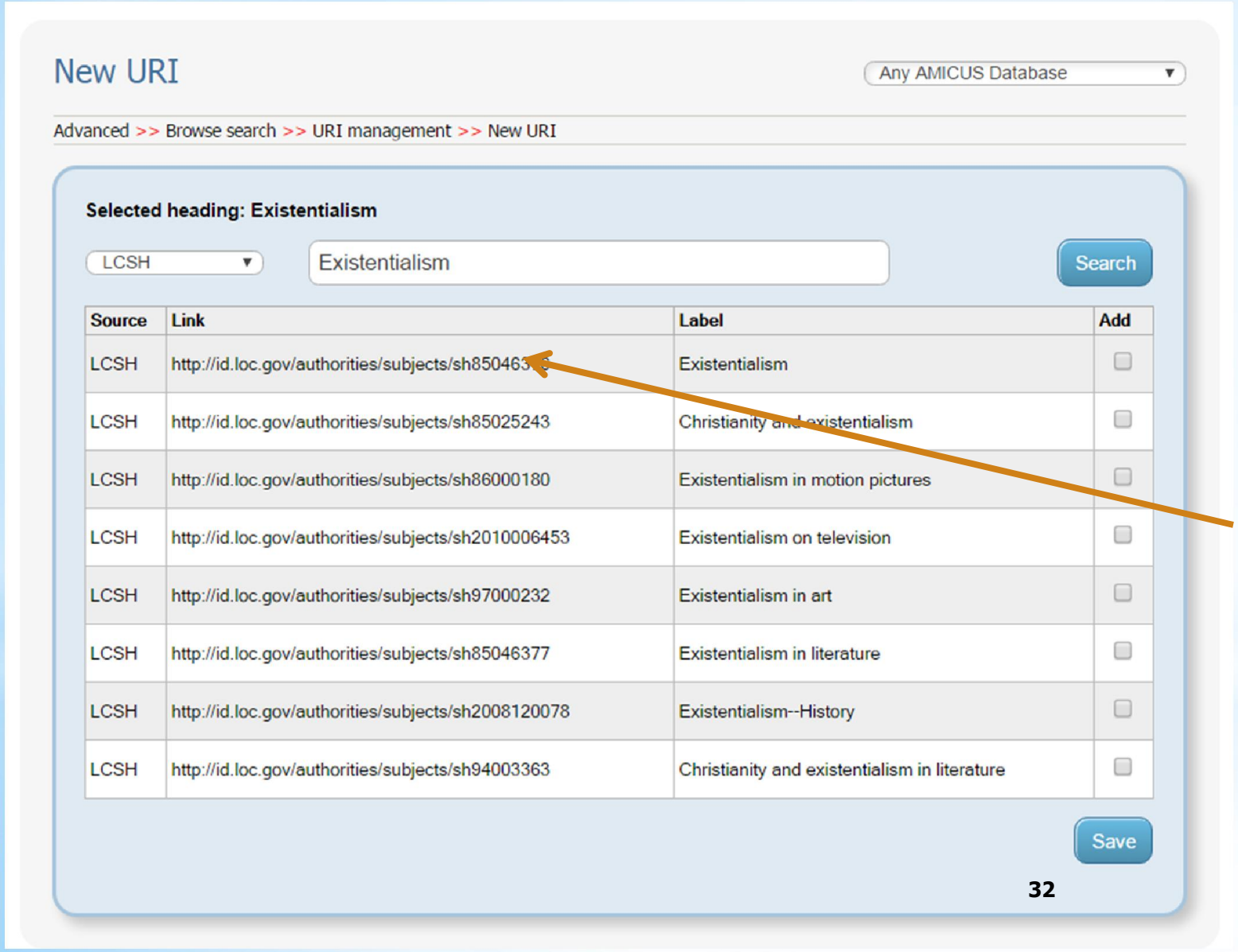
Heading	DB	URI		Refs	A	Docs	Level	Index.	Acc.
Exile (Punishment) History	B1	0		0	0	1	Unverified	und	eng
Existentialism	B1	0		0	0	3	Unverified	und	eng
Experience	B1	0		0	0	1	Unverified	und	eng
Experience (Religion) Congresses	B1	0		0	0	1	Unverified	und	eng
Expressionism (Art)	B1	0		0	0	1	Unverified	und	eng
Extinct cities Greece	B1	0		0	0	2	Unverified	und	eng
Extinct cities Italy Campania	B1	0		0	0	5	Unverified	und	eng

To add a new URI from an external source



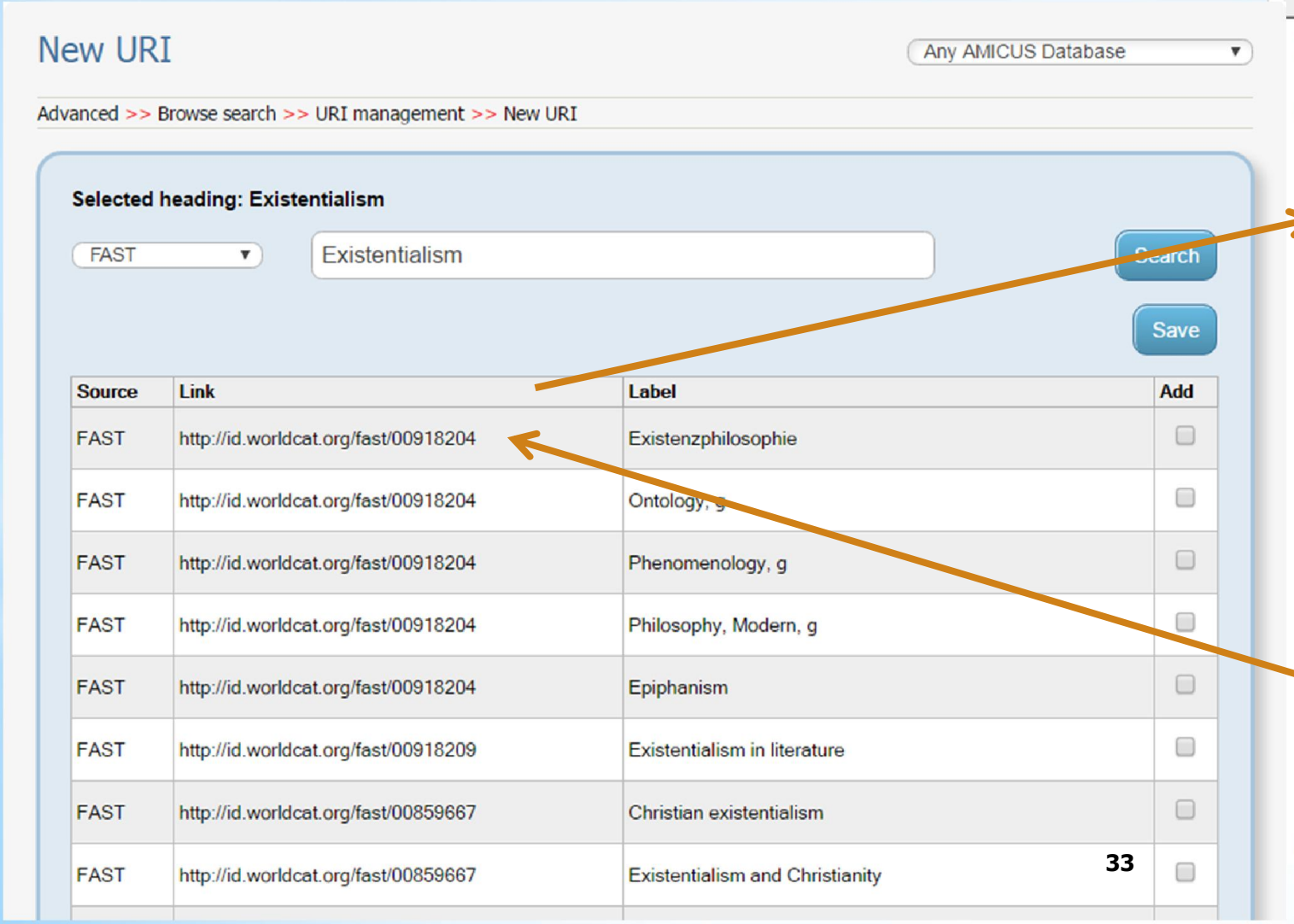
Select source to search URI

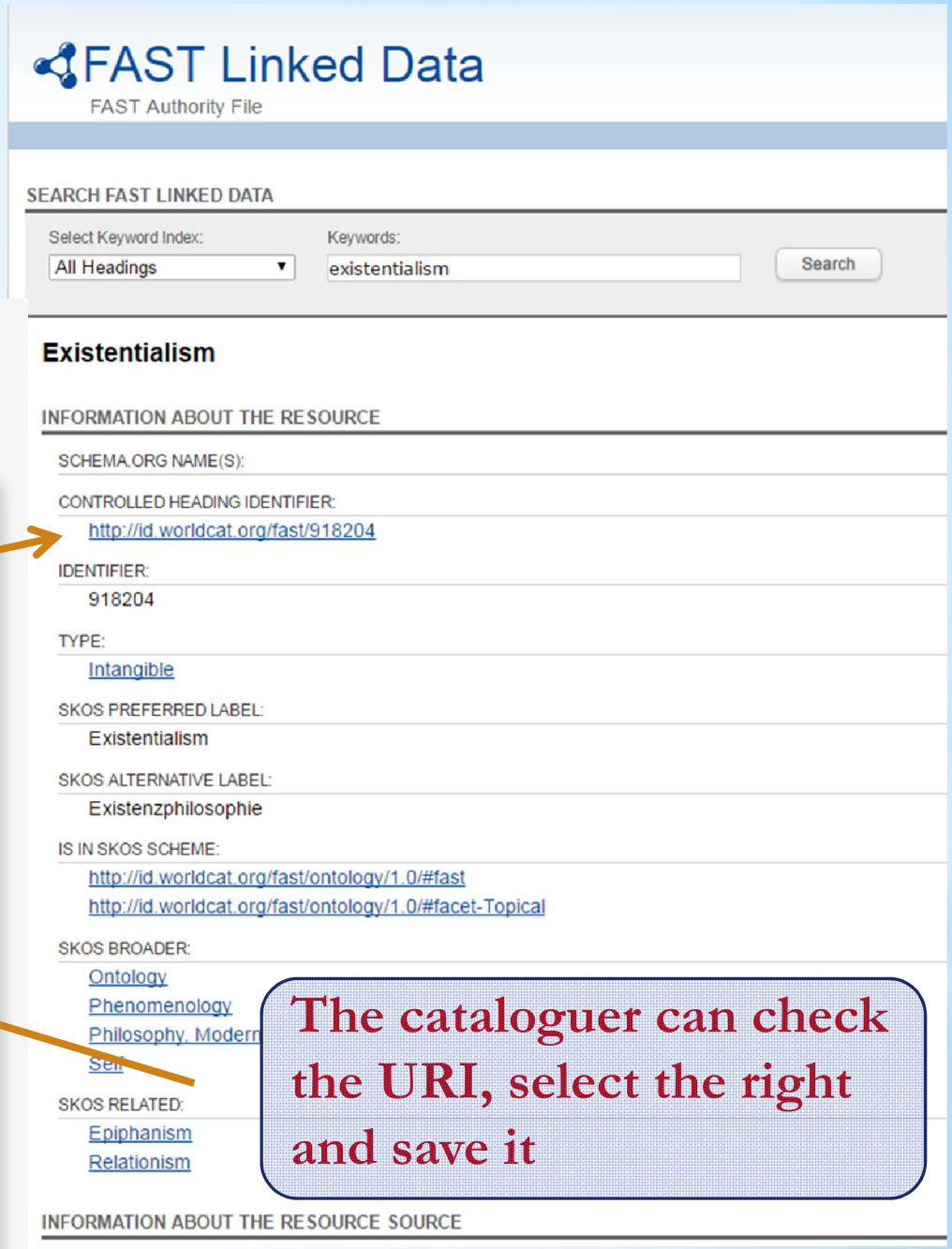
URI's result search for a source (LCSH)



Select the right URI and SAVE

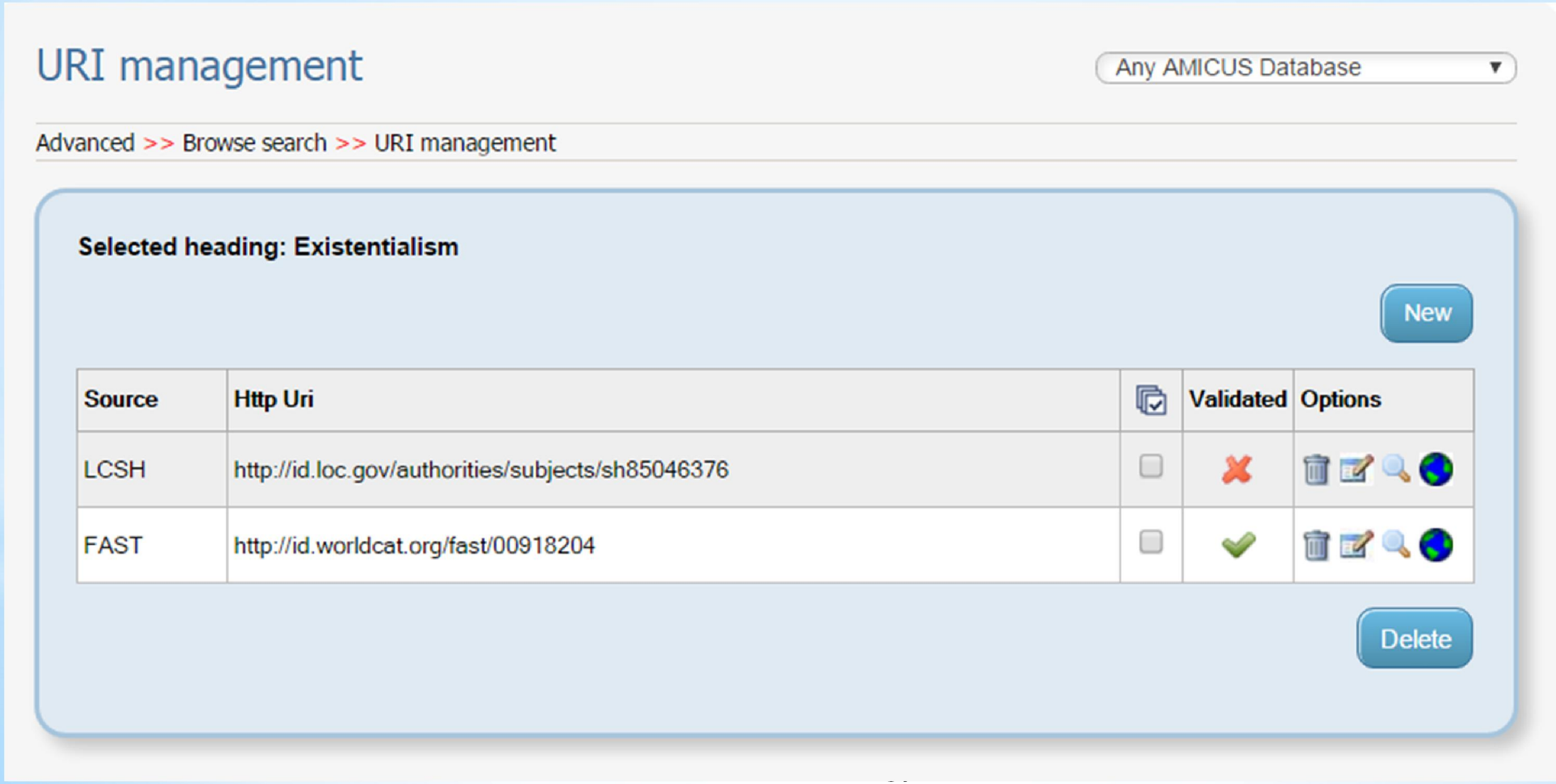
How to validate an URI: the URI quality check





The cataloguer can check the URI, select the right and save it

URI's list for the term *Existentialism*



The URI persistence → the URI Registry

Selected heading: Existentialism

New

Source	Http Uri		Validated	Options
LCSH	http://id.loc.gov/authorities/subjects/sh85046376	☐	✗	
FAST	http://id.worldcat.org/fast/00918204	☐	✓	

Delete

Any AMICUS Database

Selected heading: Existentialism

Http Uri:	http://id.worldcat.org/fast/00918204
Creation date:	28-10-2016 13:22
Creation user:	PS
Modify date:	28-10-2016 13:22
Modify user:	PS
Validated:	☑

History Save

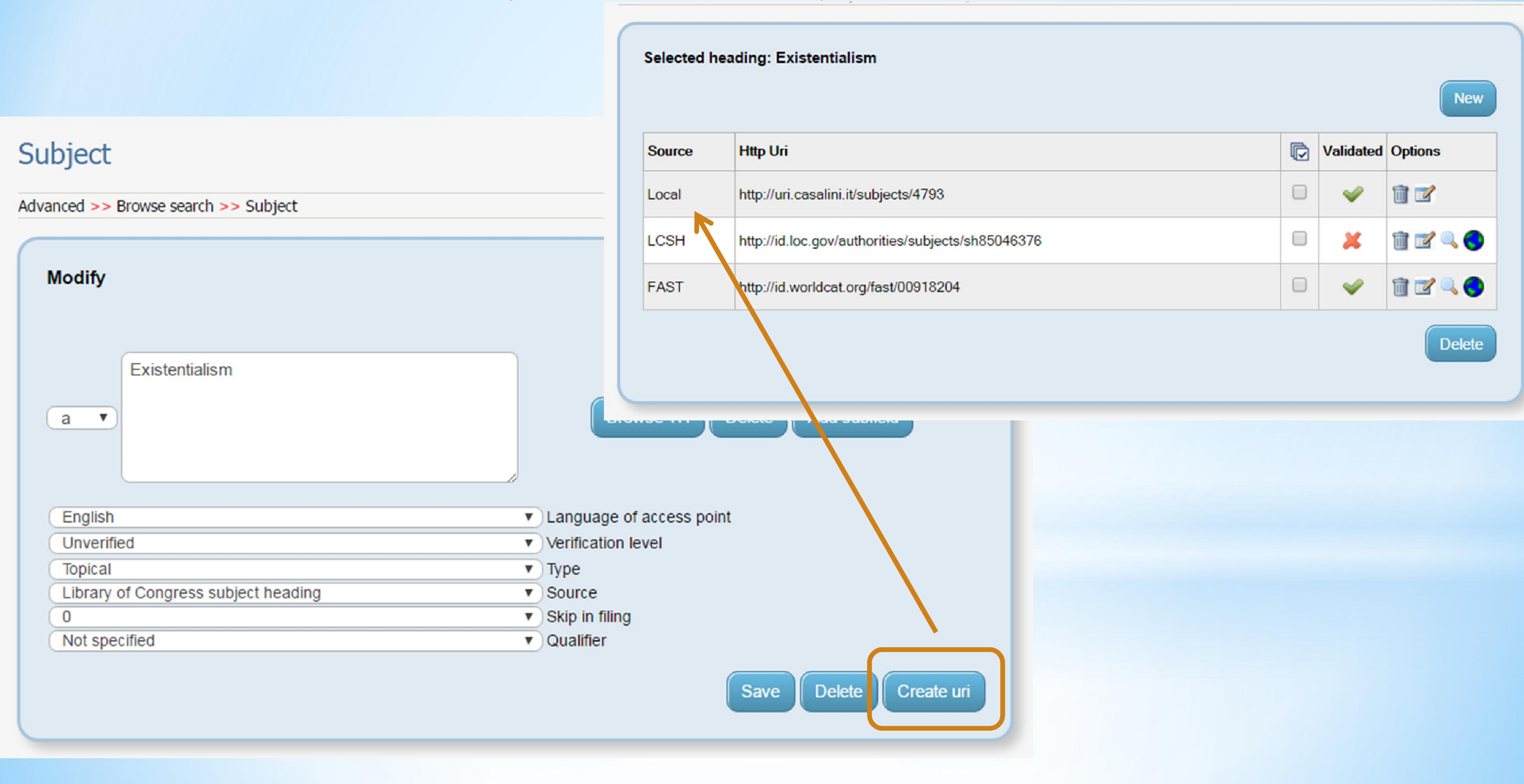
The URI persistence → the URI Registry

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

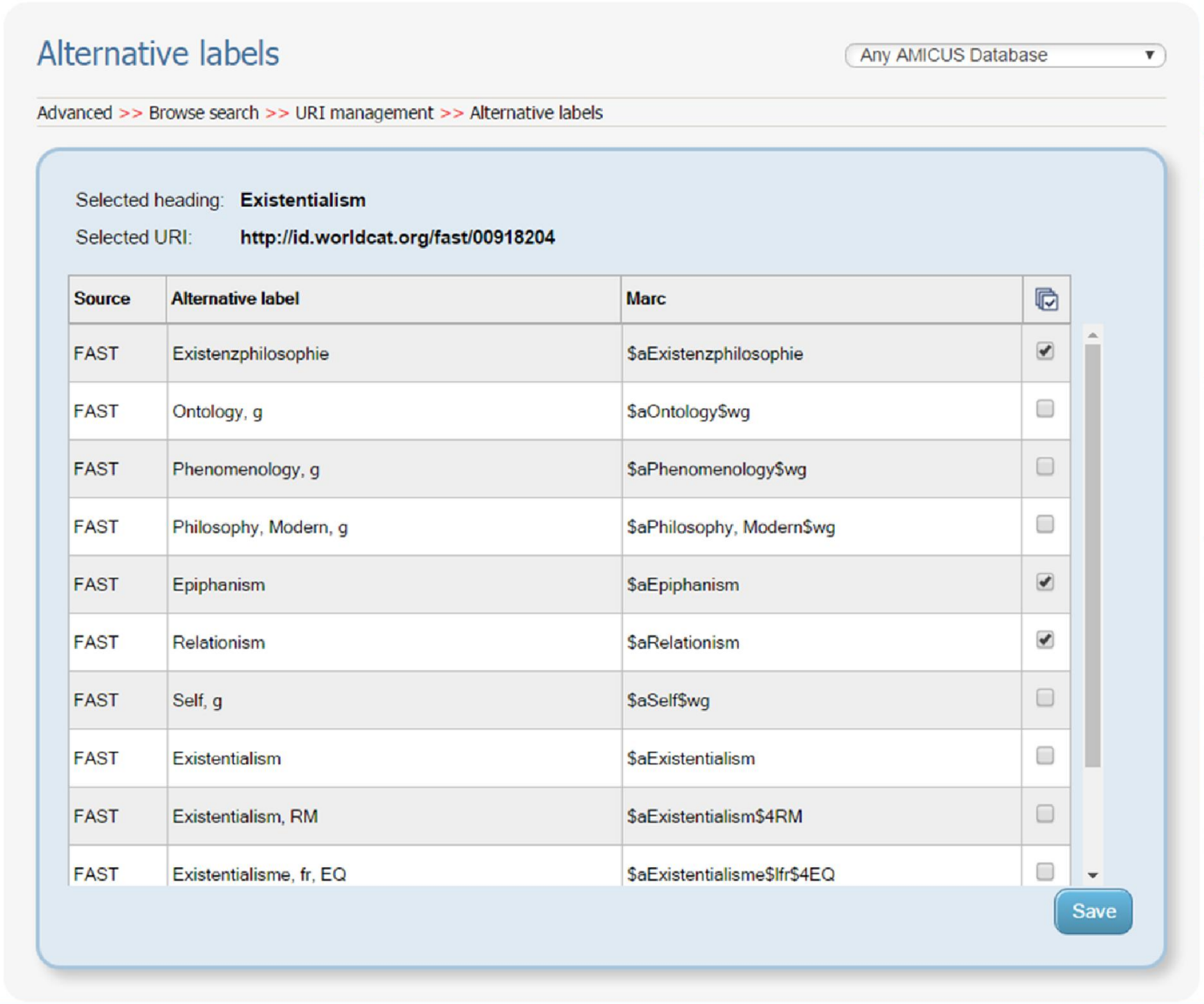
The URI Registry could keep information such as (but not limited to):

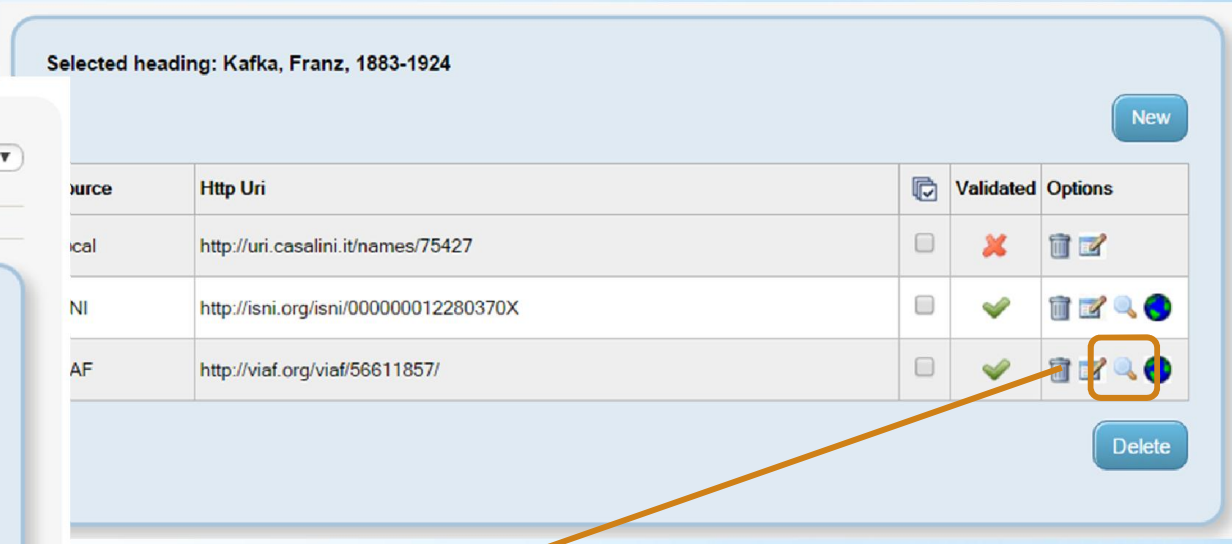
- the resources added to the cluster, but also modified or removed from it
- the date of the update
- the particular operation performed
- the status of an URI (for instance valid or invalid)
- the URI Aliases

To create a local URI (the RWO URI)



How to use URI to increase the local database (cross-references for variant forms of name)+

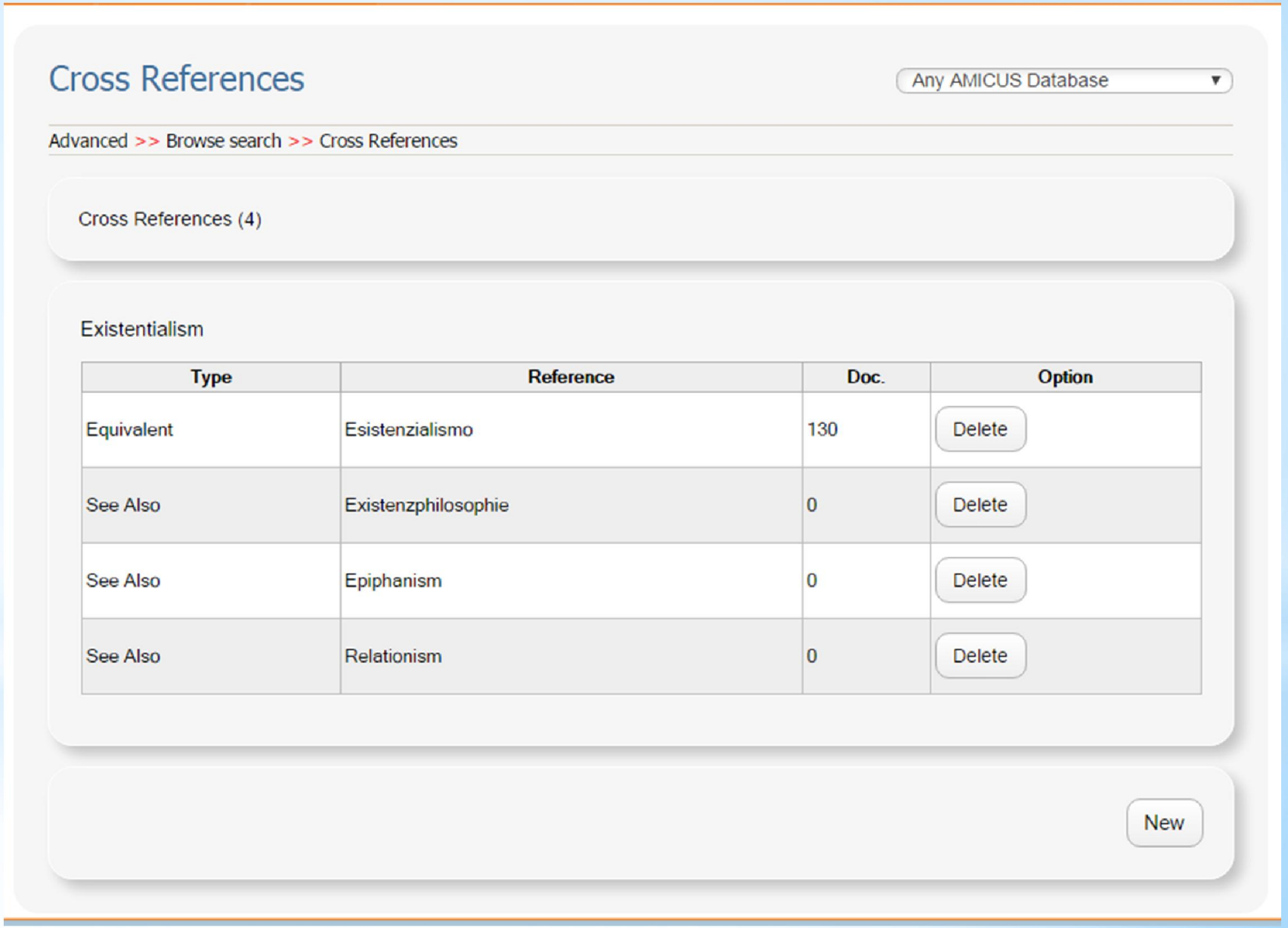




To create the cross references, select the headings from the results FAST list and SAVE

Variant forms of names coming from external sources

The cross references, coming from FAST web services now are linked to the term *Existentialism* in the Subject index



Cross References

Any AMICUS Database

Advanced >> Browse search >> Cross References

Cross References (4)

Existentialism

Type	Reference	Doc.	Option
Equivalent	Esistenzialismo	130	Delete
See Also	Existenzphilosophie	0	Delete
See Also	Epiphanism	0	Delete
See Also	Relationism	0	Delete

New

Access points and URIs

The URIs associated to a heading can be used in varying and useful ways.

In the data export/conversion process we can choose how many URIs to make available for each heading, how to associate them to the heading, how to show them in relation to data use and formats.

This export considers different customers profiles (so that each one can choose which sources to use and how to register URI)

Access point and URIs (example 1, for a Name)

As $0 associated to 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**$0(isni) 0000 0001 2280 370X.**
=245 03$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

Access point and URIs (example 2)

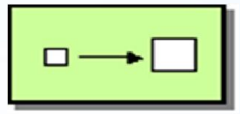
As specific tag in the MARC authority record:

=LDR 00698nz 2200145 4500
=001 000000000617
=005 20160108125155.0
=008 751003s1974\\\\enk\\\\\\\\\\\000\1\eng\\
=024 7\$a56611857$2viaf
=024 7\$a0000000122803 70X$2isni
=040 \\$aPS$bita
=100 1\$aKafka, Franz$d1883-1924
=400 1\$aKafka, F.$q(Franz)$d1883-1924
=670 \\$aWikipedia, Oct. 25, 2012$bFranz Kafka; born 3 July 1883 in Prague; died 3 June 1924 Kierling near Vienna; an influential German-language writer of novels and short stories, regarded by critics as one of the most influential authors of the 20th century. Kafka was a Modernist and heavily influenced other genres, including existentialism)

Access point and URIs (example 3)

As **RDF property** in the triples produced in the conversion process:

```
001     000000000617
024 7  $a56611857$2viaf
024 7  $a0000000012280370$2isni
100 1  $aKafka, Franz
```



```
<atcult:eb-617>
<rdf:type>
<bf:Identifier>

<atcult:eb-617>
<bf:local>
<atcult:617-kafka-franz>

<atcult:eb-617>
<bf:identifierValue>
"617"

<atcult:eb-617>
<owl:sameAs>
"http://viaf.org/viaf/56611857"

<atcult:eb-617>
<owl:sameAs>
"http://isni-url.oclc.nl/isni/
 000000012280370
```

Conclusions

The overall new approach contributes to the preconditions for a further improvement of the cooperation between institutions, the reuse of data in different scenarios, and will allow a more efficient identification of entities in the web environment, supporting, at the same time, the working process of the cultural heritage and the wider exploitation of data.

That's all

Thanks

Tiziana Possemato

tiziana.possemato@casalini.it
tiziana.possemato@atcult.it

SHARE Catalogue is online at
http://catalogo.share-cat.unina.it/sharecat/clusters?l=en

OLISuite and SHARE Catalogue are developed by @Cult