

Milan, April 2015

OPEN MEMORY PROJECT

A Linked Data Web Portal to publish and access resources on the History of the Jews and the Shoah in Italy

The Open Memory Project has been realized by the Centro di Documentazione Ebraica Contemporanea CDEC Foundation in Milan in partnership with Regesta.exe from Rome.

The CDEC Foundation (www.cdec.it) is an institute of research involved in the preservation of the cultural and historical heritage of the Jews in Italy in the contemporary age. It was born in 1955 and since then its special focus is on the documents and testimonies about the Shoah in Italy. Regesta.exe (www.regesta.com) is one of the most important Italian IT companies, specialized in LOD technologies applied to the cultural heritage.

The main aim of the Open Memory Project is the publication on the web of information and resources coming from the CDEC Foundation's archives and library, using the Linked Open Data Technologies.

Because of the huge amount of the CDEC's resources (and related topics), the first step of this wide working road-map regarded those concerning the persecution and deportation of the Jews in Italy between 1943-1945.

Shoah related resources

The CDEC's Shoah related resources and data were scattered among its different working areas - historical archive, photographic archive, audio-video archive, library, historical research area – that is, among a variety of databases, each one with its own peculiarity: inventories, catalogues, working databases realized by single researchers.

Given this basic ground, the first thing to do was the integration of all the available databases in a unique management platform in order to make possible the retrieval of data, information and digitalized documents with a unique search.

The integration process was done exploiting the linked open data technologies.

The databases have been transferred in a unique digital repository based on a Linked Data Platform (Byggle, <http://byggle.net>) that works as a unique point of access, management and publication of data and related archival and librarian resources.

The analysis of the data we were dealing with led us to take the names and biographical data of people as the red-thread of the integration process.

In particular the database of the Names of the Victims of the Shoah in Italy, due to its structure and details, appeared the most fitting to our task. In fact, it gave us the way to the “reconciliation” process and then to the “reasoning” activity.

Reconciling the Names: the Shoah domain ontology

The database of the Names of the Victims of the Shoah in Italy was the result of a multi-year research work, conducted by the CDEC since the 70's.

The database was composed of more than 19.000 record-names and more than 30 fields concerning the biographical and persecution data of each of the Victims and their relatives: name, surname, date of birth, place of birth, nationality, gender, father's name, mother's name, religion, parent's religion, place of residence, occupation, place of arrest, detention place(s), place of gathering before the deportation, date of leaving of the deportation convoy, date of arrival of the deportation convoy, convoy number, nazi-camp of arrival, eventual transfers from a nazi camp to another, serial number, fate (death, date of death /liberation, date of the return, place of the return).

The analysis of the available ontologies and the evaluation of their inadequacy to describe the persecution data, led us to create a specific domain ontology. Anyway to foster interoperability we used some of the most widespread and consolidated ontologies, such as FOAF, Biographical Vocabulary, Dublin Core, Geonames.

The relation between the two classes, "Person" and "Persecution", is the load-bearing axis of the Shoah ontology.

The "Person" class describes the biographical data and fate of each of the Victims; the "Persecution" class describes every single phase of the persecution and deportation process - from the arrest in Italy to the nazi-camp internment (or selection and death just after the arrival, in many cases) - and it represents the core and the real innovation side of the Shoah Ontology (see the reference document, <http://dati.cdec.it/iod/shoah/reference-document.html>). The step after has been the assignment of a persistent IRI to every piece of information coming from the original database.

Then, the data about the Victims has been reconciled, also with data coming from other databases (the database of the antifascist and partisan Jews, for example, converted in RDF/xml, not published yet): names with names, places with places, and then, names with places¹ (birth places, residence places, arrest places, detention places) with convoys, with nazi-camps.

In this way the previous scattered data are now semantically linked in a network of relations concerning people and families, places, persecution experiences.

Families members have been virtually rejoined, as well as people coming from the same places, or gathered and deported from the same place on the same convoy in the same days to the same nazi-camp; or people murdered before the deportation, during the same nazi-fascist massacre (for ex. the Fosse Ardeatine massacre, the 24 March 1944: 335 people, included 76 Jews, were killed in retaliation for the anti-nazi attack in Rome the day before).

Data has been also interlinked through the "same as" relation to the available sources coming from the LOD Cloud (DBpedia, VIAF, Geonames, dati.camera, dati.acs) and the data set included in the LOD cloud itself (last updated September 2014).

¹ For example, the place of Rome is linked to 1688 persons, 1755 persecution experiences, 14 antifascist activities, 17 partisan activities, 1 massacre.

The data set of the Victims of the Shoah in Italy has been published in a Sparql endpoint where data can be queried, downloaded and reused as well as interlinked to other data sets (under CC share-a-like licence).

You can also browse the data through the LOD browser LODLIVE (<http://dati.cdec.it/lodlive/?http://dati.cdec.it/lod/shoah/person/4919>) and the new LODVIEW (<http://dati.cdec.it/lod/shoah/person/5002>)

The reasoning activity: connecting data to archival and cataloguing descriptions, and digital documents. The role of the linked data platform.

The RDF/xml conversion and the assignment of the IRIs, has been the crucial step to the second phase of our project: the connection of every entity to the librarian and archival descriptions and the (eventual) digital documents.

In this task the linked open data platform Bygle has played a key-role: it worked as a unique point of access, management and connection between the RDF/xml datasets and the CDEC inventories and catalogues located in the OS xml platform xDams (xdams.org).

The traditional CFP authority files (as well as the places) expected in the inventories and catalogues made according to the ISAAD/ ISAAR standards, have been “replaced” by the IRIs of the persons (and places).

The reasoning activity made possible the real integration of the CDEC’s documental and information heritage: the descriptions of papers, photographs, audio-video recording as well as the related digital resources (when available) concerning a specific person (a Victim of the Shoah, in this case) have been linked to that “person” and are now available through a unique search.

For ex., the person entity “Levi, Primo” (<http://dati.cdec.it/lod/shoah/person/5002>) is currently linked to:

Audio-Video Archive (2 occurrences)
Historical Archive:
Antifascist and Partisan Jews Fond (3 occurrences)
“Vicissitudini dei singoli” Fond (8 occurrences)
“Comitato Ricerche Deportati Ebrei” Fond (1 occurrence)
“Massimo Adolfo Vitale” Fond (3 occurrences)
Photographic Archive (42 occurrences)

“Levi, Primo” is also currently connected to the library resources, that is 63 monographies and 4 articles.

From “inside” to “outside”.
The CDEC Digital Library, a Linked Open Data Web Portal

We mentioned before the key-role played by the LOD Platform Bygle in the reasoning activity.

This platform has been crucial also in the third and last phase of our project: opening our archives and make our resources available to the web users.

The CDEC Digital Library web portal (<http://digital-library.cdec.it>) is the innovative output of this task.

In fact, despite the traditional graphic interface, this web portal finds its basic structure in the LOD platform - instead of the usual network of databases and multiple queries.

Thanks to this choice, through the CDEC Digital Library we are able to provide plenty of information and resources to the users: those coming from our archives and library and those coming from the LOD Cloud; furthermore, you can also download the RDF data of each Victim.

The CDEC Digital Library is one of the first cultural web portals that display data concerning its own heritage together with data coming from the LOD cloud (<http://digital-library.cdec.it/cdec-web/person/detail/person-5002/levi-primo.html>).

The challenge for the future

Our specific challenge in the next future is the completion of our road map and in particular the publication, in linked open data format, of the more than 13.000 names of the rescued Jews and their rescuers. With this task we will be able to virtually rejoin almost the 80% of the Jewish families persecuted in Italy from 1943 and 1945.

In a more general perspective what we expect from the future is the widespread of the LOD technologies among the Jewish and non Jewish institutes and museum, and in particular among all of them involved in the research and publication of the names of the Victims of the Shoah and related resources. This would be a fundamental step toward the reconciliation of names currently listed in a variety of databases, sometimes with different spellings, sometimes with light variants of the biographical data².

Our hope for the future is to get a IRI for each of the names of the Victims of the Shoah in Europe. To this task the way has been already outlined by and thanks to the Shoah ontology that we built and that is open to be expanded by descriptions coming from other experiences of

² A pretty clear example regards the name of a polish deportee “Jeruchem Flank”. This name is currently present in four different databases: the CDEC one (<http://digital-library.cdec.it/cdec-web/person/detail/person-2634/flank-jeruchem.html>), the Memorial de la Shoah in Paris (<http://bdi.memorialdelashoah.org/internet/jsp/core/MmsRedirector.jsp?id=11494&type=VICTIM>), the Yad Vashem in Jerusalem (<http://db.yadvashem.org/names/nameResults.html?lastName=flank&lastNameType=THESAURUS&firstName=jeruchem&firstNameType=THESAURUS&language=en>), and the Documentation Center of Austrian Resistance - DOEW in Wien (http://www.doew.at/personensuche?gestapo=on&findall=&lang=en&shoah=on&politisch=on&spiegelgrund=on&firstName=jeruchem+%&lastName=flank&birthdate=&birthdate_to=&birthplace=&residence=&newsearch=10&iSortCol_0=1&sSortDir_0=asc&lang=en&suchen=Ricerca#). Maybe the same name is also listed in other databases that we didn't find or we don't know. Every one of the mentioned databases provide complementary information: the CDEC provides the name of Jeruchem Flank's wife and the exact deportation iter; the Memorial de la Shoah provides variants about the personal name (Jeruchem Plank) and the place of birth (Dobromil; Dohomil) and information about his occupation and the number of the deportation convoy from Drancy to Auschwitz. Yad Vashem provides a further variant of the personal name and biographical information similar (but not identical) to those provided by the CDEC and the Memorial de la Shoah. The DOEW database provides basic information (name, surname, place and date of birth, deportation iter, fate) already given in the other databases.

Open Memory Project
Laura Brazzo, Silvia Mazzini

persecution such as, for example, those concerning the east-European Jews (where the reclusion in the ghettos represents a specific and typical phase of the deportation process).

A work similar to our one, made on a large scale would improve the retrieval of documents and information and, consequently, would help and facilitate the work of scholars and researchers.