

THE ADOCHS PROJECT PRESENTS

# LINKING THE PAST

Possibilities and limits of publishing  
archival data as Linked Data

November 22, 2019

9AM - 5PM

KBR, BRUSSELS

FREE REGISTRATION AND PROGRAM:  
[WWW.ADOCHS.BE/LINKING](http://WWW.ADOCHS.BE/LINKING)

---



## LINKING THE PAST | 22 11 2019 | KBR

### Possibilities and limits of publishing archival data as Linked Data

- 09:00        **Reception**
- 09:30        **Introduction**  
Seth van Hooland (Université Libre de Bruxelles)

#### Session 1: New Data Models

- 10:00        **Records in Contexts (RiC) conceptual model and ontology**  
Florence Clavaud (Archives nationales of France and International Council on Archives)
- 10:30        **The Matterhorn RDF Data Model**  
Tobias Wildi (Docuteam)
- 11:00        Coffee break
- 11:30        **Wikidata and Archival Metadata: opportunities and challenges for heritage institutions**  
Baptiste de Coulon (Le lieu imaginaire)
- 12:00        **IIF protocol: a standard to publish and easily share the digital pictures.**  
Jean-François Moufflet (National Archives of France)
- 12:30        **Questions**
- 13:00        Lunch break

#### Session 2: Linked Open Data – World War II Case Studies

- 14:00        **WarSampo – Finnish WW2 on the Semantic Web**  
Eero Hyvonen (University of Helsinki, Helsinki Centre for Digital Humanities – HELDIG and Aalto University, Department of Computer Science, Semantic Computing Research Group – SeCo)
- 14:30        **From CegeSoma spreadsheets to Linked Data: a Wikibase journey**  
Anne Chardonens (Université Libre de Bruxelles - ReSIC)
- 15:00        Coffee break
- 15:30        **Describing the Persecution of the Jews in Italy: the OWL Domain “Shoah Ontology”**  
Laura Brazzo (The Foundation Jewish Contemporary Documentation Center)
- 16:00        **Debate**
- 16:30        **Conclusion: The new data models, a real opportunity?**  
Frédéric Lemmers (The Royal Library of Belgium – KBR)
- 17:00        **End**

## Records in Contexts (RiC) conceptual model and ontology

Florence Clavaud (Archives nationales of France and International Council on Archives)

The Expert Group on Archival Description of the International Council on Archives (ICA-EGAD) is developing a generic, international, standard on archival description. This global framework will replace the four ICA existing standards (ISAD(G), ISAAR(CPF), ISDF and ISDAH) that were published between 1994 and 2008 and are now being used by the organizations and persons in charge of archival fonds; it will be based on the definition of the entities that together constitute the world of archives. Moreover, it aims to represent, in a clearer and more complete way than the previous standards, the layers of contexts in which records and record sets exist, from their creation to the events that may affect them once transferred in an archival repository. The target is to enable a multidimensional archival description, that is more accurate, that can be more easily accessed and understood by human users and that can also be more finely processed by the machines, than what you get when you use the four existing metadata standards.

This new standard, named Records in Contexts (RiC), will consist of three parts: a global, entity-relationship, conceptual model (which will define the archival entities, their attributes and the relations that connect them to other entities); an ontology (i.e. a formal, OWL, representation of the conceptual model, that can in particular be used for creating and publishing online, as Linked Data, graphs of archival metadata expressed in RDF); and application guidelines.

ICA-EGAD published a 0.1 version of RiC Conceptual Model (RiC-CM) in September 2016. A preview of RiC-CM v0.2, resulting from a significant amount of work (assessing the previous version, increasing its accuracy and changing its form) should be released by the end of November 2019. At the same time, a 0.1 version of RiC-Ontology (RiC-O) should also be published. Both releases will go along with a public call for comments.

The presentation will give an overview of RiC-CM v0.2 and of the design principles of RiC-O. It will also show some concrete examples of RDF datasets conforming to RiC-O. From the projects that have been undertaken at the Archives nationales de France for two years, it will discuss the paths that RiC opens for archivists, records managers and users, including researchers, the issues that have to be solved and the ways that can be followed when implementing it, and, more generally speaking, the nature of the evolution that is just beginning.

*Florence Clavaud is archiviste-paléographe and conservateur en chef du patrimoine. She has been the head of the Office on Authority Records and Vocabularies at the Archives nationales de France since 2013 and was, from 1990 to 2004, successively head of its Records Reproduction and Information Technologies services. She worked in the private sector, then as a lecturer and programme director of the History and new technologies master's degree at the École nationale des Chartes, where she remains a member of the research team, actively involved in several digital humanities projects.*

*She is a member of the Technical Subcommittee on Encoded Archival Standards of the Society of American Archivists, and an executive member of the International Council on Archives Expert Group on Archival Description (ICA-EGAD), responsible for Records in Contexts Ontology. More details here: <http://www.chartes.psl.eu/fr/florence-clavaud>.*

## The Matterhorn RDF Data Model

Tobias Wildi (Docuteam)

Matterhorn RDF is a linked data-based model for archival metadata with the goal of improving the contextualization of archival records. It covers the three standards ISAD(G), ISAAR(CPF) and ISDF, as well as the areas "Preservation Description Information" and "Representation Information" of the OAIS information model. For the implementation of Matterhorn RDF, classes and properties of existing ontologies are used. The formalization of the model is done with the help of SHACL shapes.

Matterhorn RDF is not a competitor to Records in Context, but wants to depict the RiC concept model as soon as a stable version is available. The goal will be to implement the RiC concept model with existing ontologies.

Matterhorn RDF is a joint development of the "Archives de l'Etat du Valais" (Alain Dubois) and docuteam (Tobias Wildi). Beginning of 2019 a broader group of Swiss archives, including the Swiss National Archives started to develop a linked data model for archives based on the RiC conceptual model. Matterhorn RDF is a preparatory work for this new national model.

*Tobias Wildi is director of docuteam, a Swiss company that supports a wide range of organizations with their information management and archiving. Tobias holds a Master in History and IT, a PhD in History and a Certificate in archival and information science. He primarily works in projects related to digital long-term archiving and the protection of digital cultural assets. Tobias is the president of the Swiss Federal Commission for the Protection of Cultural Property and since June 2019 member of the International Council on Archives Experts Group on Archival Description (EGAD).*

## Wikidata and Archival Metadata: opportunities and challenges for heritage institutions

Baptiste de Coulon (le lieu imaginaire)

The knowledge base Wikidata is a collaboratively edited online centralized repository, the use of which is free. The data are structured according to Resource Description Framework (RDF) specifications and published under the CC0 free licence. Wikidata do not store original information, all statements (entity–attribute–value) must have references. From our point of view, Wikidata will be the next key tool for heritage institution's professionals, if it is not already. We see three main uses : find information, clean and link authorities, give visibility to their collections. In our intervention, we will explore these three dimensions, but also insist on the first-hand necessity for institutions in a LOD perspective to maintain persistent identifiers on their online data.

*Baptiste de Coulon is an archivist and information manager in Switzerland. He works at his own company called « le lieu imaginaire » and is at the same time employed by the Library of the University of Lausanne as metadata manager in the Manuscript's Department. He has been conducting many projects for heritage institutions and societies for over ten years. He is involved in introducing a large use of the Wikimedia platforms in the daily job of the patrimonial institution's professionals.*

## The IIIF protocol: How to link images ?

Jean-François Moufflet (National Archives of France)

Much effort has been made by cultural institutions to standardize the descriptive metadata, aiming to homogenize their structure and improve their interoperability for their publication. Today, with the influence of the Semantic Web, we tend to link these metadata sets to create much richer knowledge networks and benefit from the complementarity of the different repositories.

The clustering of cultural resources around shared portals concerns also digitized documents, whether they are manuscripts, books and of course archives. Whereas the digital pictures of the documents were produced and managed by different institutions, many projects in which these institutions are involved today aim to aggregate the digital contents in a common consultation interface.

In other words, we may need to standardize the dissemination of the pictures if we want to make their aggregation easier. This is what the IIIF (International Interoperability Image Framework) protocol was developed for. It is an international technical framework, providing the institutions with norms to share their pictures among common websites.

The IIIF protocol comes with APIs that not only call pictures stored separately, but also can manipulate them with great flexibility. Both the pictures and their metadata can be displayed with many possibilities. One can change on-the-fly the size, the colors or the direction of the picture. Other pictures can be overlaid on the first one or they can be displayed to be compared. Textual content resulting from an optical character recognition or from collaborative tagging can be displayed and searched in the picture...

Commonly used by the digital libraries in European sized projects, the IIIF protocol still remains more discretely used by the archives services. The purpose of this presentation is to show its great potential.

*Jean-François Moufflet, "archivist paleographer" (former student of the École des Chartes), chief curator, responsible of archival fonds in the French national Archives. Jean-François Moufflet first worked at the "Archives de France", where he had to advise the French archives services network for the topics related to digitization, software and digital preservation, and then in the "Institut national du patrimoine" as responsible for the initial training of the curator profession. Today, he is responsible of archival fonds in the French national archives and is devoted to link archival processes to technological innovations. He recently participated in the HIMANIS project, where was experimented handwritten text recognition, whose results were published on a portal using the IIIF protocol.*



## WarSampo – Finnish WW2 on the Semantic Web

Eero Hyvonen (University of Helsinki, Helsinki Centre for Digital Humanities – HELDIG and Aalto University, Department of Computer Science, Semantic Computing Research Group – SeCo)

The talk gives an overview of the project “WarSampo - Finnish WW2 on the Semantic Web”. WarSampo is based on harmonizing massive distributed datasets using event-based modeling, which makes it possible to enrich the datasets semantically with each other’s’ contents. WarSampo has two components: First, a Linked Open Data (LOD) service WarSampo Data for Digital Humanities (DH) research and for creating applications related to war history. Second, a semantic WarSampo Portal has been created to test and demonstrate the usability of the data service. The WarSampo Portal allows both historians and laymen to study war history and destinies of their family members in the war from different interlinked application perspectives. Published in November 2015, the WarSampo has been extend in 2017 with a new application perspective on military cemeteries and in November 2019 a new application on Finnish prisoners of War in the Soviet Union 1939-1945 will be released. In 2017, WarSampo got the international LODLAM Open Data Prize in Venice. The portal had in 2018 230 000 distinct users showing a great public interest in applications like this.

Project homepage with videos and publications online:

<https://seco.cs.aalto.fi/projects/sotasampo/en/>

The semantic portal in use: <http://sotasampo.fi/en>

*Eero Hyvönen is director of Helsinki Centre for Digital Humanities (HELDIG) at the University of Helsinki and professor of semantic media technology at the Aalto University, Department of Computer Science, where he directs the Semantic Computing Research Group (SeCo) specializing on Semantic Web technologies and applications. A major recent theme (since 2001) in his research has been the development of the national level semantic web infrastructure and its application in Digital Humanities and other areas. Eero Hyvönen has published over 400 research articles and books and has got several international and national awards. He acts in the editorial boards of Semantic Web - Interoperability, Usability, Applicability, Semantic Computing, International Journal of Metadata, Semantics, and Ontologies, and International Journal on Semantic Web and Information Systems, and has co-chaired and acted in the programme committees of tens of major conferences. <https://seco.cs.aalto.fi/u/eahyvone/>*

## From spreadsheets to Linked Data: a Wikibase journey

Anne Chardonnes (Université Libre de Bruxelles - ReSIC)

The web of data leads to an evolution of galleries, libraries, archives, and museums (GLAM) authority data, which gradually move from strings to entities identified by Uniform Resource Identifiers (URIs). These URIs help to go beyond the pale of natural language ambiguities, but also to link these entities to others in a machine-readable format. However, this transition is not trivial to implement. How can an institution with limited means consider sharing and linking data while still using local tabular files to store information?

I investigate this question through a case study based on the nominative data of CegeSoma (Centre for Historical Research and Documentation on War and Society) and the use of Wikibase, the free software behind Wikidata. Besides enabling integration of heterogeneous sources in the same data infrastructure, Wikibase can also assign persistent identifiers to individuals, deal with multilingual content, and query data through SPARQL queries. This talk aims to show the opportunities and challenges of this process, from name disambiguation to data modelling, with an emphasis on the interaction between Wikidata and Wikibase.

*Anne Chardonnes holds a Master's degree in Cultural Management from the Université Libre de Bruxelles (ULB). She is currently pursuing a PhD in Information and Communication Sciences and Technologies at the same university. Since November 2016, she has joined CegeSoma to take part in the ADOCHS project, which aims to improve quality control processes in the context of cultural heritage digitization. Her work focuses on user needs, participatory practices, and the impact of web data on archival (meta)data.*



## Describing the Persecution of the Jews in Italy: the OWL Domain “Shoah Ontology”

Laura Brazzo (The Foundation Jewish Contemporary Documentation Center)

This presentation will focus on the development of the OWL domain Shoah Ontology, carried out at the Cdec Foundation and devoted to the formal description of the persecution of the Jews in Italy during the Nazi occupation (1943-1945) – from the arrests the detention places to the deportation towards the concentration and extermination camps.

The presentation will explain the reasons why we opted for developing a domain ontology instead of reusing already existing ontologies. It will also illustrate the classes and properties of the Shoah Ontology as well as the benefits of the Linked Data solution in terms of knowledge enrichment for both the promoting institution and the audience.

*Laura Brazzo received her degree in Contemporary History at the Università Statale of Milan. Her thesis concerned the Italian Zionists and the Italian diplomats dealing with the question of the future of Palestine during the First World War. Her thesis was in Political science and concerned with the relations between the EEC and the State of Israel and the construction of the "image" of Israel by the European Institutions from 1957 to 1973. For almost ten years she is involved in the activities of the Chair of Contemporary History at the Università Statale in Milan and, at the same time, with the Commission of History of International Relations. She worked as an archivist at the Italian Institute for International Political Studies in Milan and then at the Fondazione Centro di Documentazione Ebraica Contemporanea ([CDEC](#)) where she is since 2005.*