Fraunhofer
IBMT

FRAUNHOFER-INSTITUT FÜR BIOMEDIZINISCHE TECHNIK IBMT

1 *Paper-based archiving in the past.*
2 *Clinical digital preservation today and its problem of obsolescence of media types, data formats, metadata standards and terminologies.*

An Ontology Framework for Long-term Digital Preservation and beyond

Fraunhofer Institute for Biomedical Engineering IBMT
Prof. Dr. Heiko Zimmermann
Prof. Dr. Günter R. Fuhr
Joseph-von-Fraunhofer-Weg 1
66280 Sulzbach
Germany

Contact

Health Information Systems
Dipl.-Inform. Stephan Kiefer
Telephone +49 6894 980-156
Fax +49 6894 980-400
stephan.kiefer@ibmt.fraunhofer.de

www.ibmt.fraunhofer.de

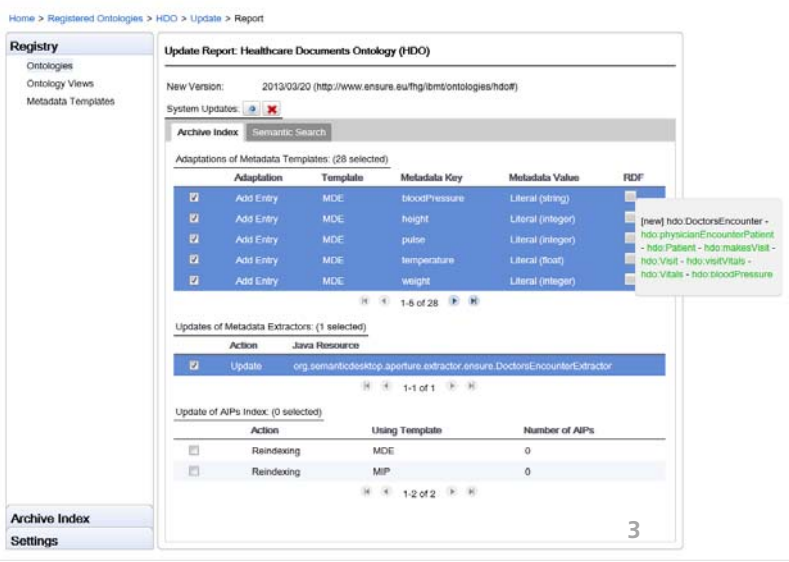
ENSURE

The terminology problem in long-term preservation

Healthcare organisations and clinical researchers are producing an exponentially growing amount of digital born data while they are confronted with the problem to maintain these sensitive mass data for future use. Long-term preservation of personal electronic health information is a key issue for life-long electronic health records, however, it is poorly implemented in healthcare institutions and little attention is given to problems like obsolescence of formats and EHR applications or changing regulations and medical terminologies over the time which jeopardize retrieval and reusability of information after decades of preservation. As a result, future users may not be able to query the preserved data in their actual domain language and according to actual search criteria but in obsolete terminologies only and with old and limited search criteria. In the positive case that new terminology and search criteria can be applied, the system may be unable to retrieve data following old terminologies. A prominent example for this problem represents the evolving ICD terminology of WHO for the classification of diseases which is widely used in health information systems.

The solution: Object indexing with domain ontologies and preservation tools for ontologies management

In order to address the above preservation problem, Fraunhofer IBMT has developed an ontology-based data indexing and retrieval solution together with a tool to manage ontology and terminology evolution over the time as part of the ENSURE long-term preservation system. »ENSURE – Enabling kNowledge Sustainability, Usabil-



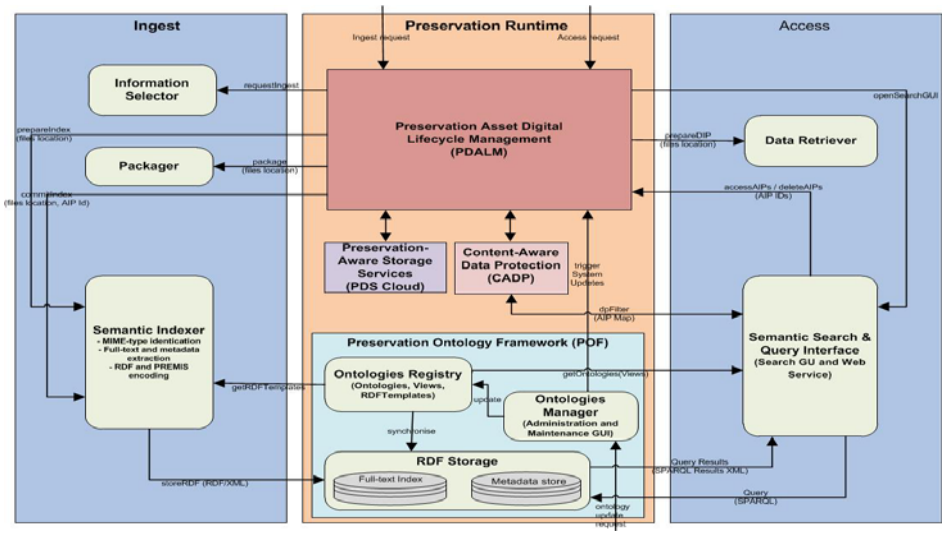
ity and Recovery for Economic Value« is an FP7 funded research collaboration of 11 European and two Israeli partners to contribute ICT innovations to the state-of-art in long-term digital preservation.

In our solution, semantic web technologies are applied to effectively model, collect and manage the metadata of the digital objects from the different domains and to provide a powerful semantic search and access mechanism to preserved data. Therefore we represent metadata of data objects in terms of an integrated set of formal ontologies. This allows modelling the preservation knowledge and domain-specific object formats and concepts in an application-oriented way through formal ontologies and leveraging them in managing the archive. The ontologies contain concepts describing general features of data objects (i. e., type, format, size, Preservation Description Information) as well as domain-specific information. The metadata of data objects are extracted by a semantic indexer component which encodes them as RDF triples that are stored by the ENSURE Preservation Runtime in an index. The ontologies together with XML metadata templates are registered as semantic resources in our so-called Preservation Ontology Framework (POF) which is part of the Preservation Runtime System. Due to its Ontologies Manager component POF also allows addressing more preservation-related problems like the evolution of ontologies or changing access policies.The Ontology Manager enables the user to maintain ontologies with their different versions over their lifecycle through a GUI (Figure 3). As a key functionality the component allows updating an ontology version and executing the required system adaptations. In some cases, the update of an ontology version in the system may require actions in the archive system, such as re-indexing of archived AIPs in order to keep their index and the entire archive system consistent. The calculation of the differences between sequential versions of ontologies, for which we applied COnto-Diff algorithm developed by University Leipzig, is used to conclude the required system adaptations, to inform the administrator respectively and to conduct these adaptations automatically where possible.

In this way the POF supports solving problems resulting from evolving domain terminologies and changing search criteria over the time.

Usage beyond preservation

The Ontologies Framework in combination with the semantic indexing and search and retrieval components can be utilised in every modern information system architecture where ontologies are used and need to be maintained.



Information preparation architecture with Preservation Ontology Framework.

Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 270000 - ENSURE
Project website:
www.ensure-fp7.eu

3 *Screenshot of Ontologies Manager.*

4 *Biobanking and clinical research -domains with long-term preservation requirements.*