

View other issues

Contents: ERCIM News 80

Knowledge Management for Digital Preservation

by Yannis Tzitzikas and Vassilis Christophides

We can preserve the bits, but what about the knowledge encoded in them? Modern societies and economies are increasingly dependent on a deluge of information that is only available in digital form. The preservation of this information in an unstable and rapidly evolving technological (and social) environment is a challenging problem of great importance. The CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) project built a pioneering framework to support the end-to-end preservation 'life cycle' for scientific, artistic and cultural information based on existing and emerging standards. CASPAR aimed to preserve not simply the bits of digital objects but also the information and knowledge that is encoded in these objects.

The key contributions of CASPAR regarding knowledge management revolve around four main topics:

(a) **Intelligibility.** Since it is hard to define explicitly what information or what knowledge is, it is equally difficult to claim that a particular approach, methodology or technique can indeed preserve information and knowledge. To tackle this issue and for preserving the meaning of digital objects, CASPAR formalized the notion of intelligibility in an OAIS-compliant manner and provided guidelines, methodologies and components that can aid humans in preserving information and knowledge. Specifically, it formalized the notion of intelligibility and an intelligibility gap through the notion of dependency. This perspective allows us to answer questions relating to (a) what kind of (and how much) representation information we need, (b) how this depends on the designated community, and (c) what kind of automation we can offer (regarding packaging and dissemination). Apart from developing formal and conceptual models, CASPAR has developed tools and applied them to real data.

(b) **Semantic Web evolution management.** Evolution is a key concept, because preservation is a dynamic process; the world evolves, software and hardware evolve, metadata schemas evolve, digital objects evolve, community knowledge evolves. This change poses several challenging requirements for semantic Web repositories, including bulk metadata updates, versioning metadata and ontologies, ontology evolution, comparison operators and change log management. CASPAR has developed an advanced platform for semantic Web management (SWKM) for tackling the above requirements.

(c) **Provenance modelling and querying.** There is a need for a comprehensive and extensible conceptual framework that will allow provenance information to be integrated, exchanged and exploited within or across digital archives. CASPAR extended the ISO standard CIDOC CRM, defining CIDOC CRM Digital to explicitly model digital objects, and showed how it can be employed for provenance queries.

(d) **Automating the ingestion of metadata.** The creation and maintenance of metadata is a laborious task whose benefits may only be obvious in the longer term. There is a need for tools that automate as much as possible the creation and curation of preservation metadata. CASPAR developed PreScan, a tool for automating the ingestion phase. It can bind together automatically extracted embedded metadata with manually provided metadata and dependency management services (recall Intelligibility), and transforms the metadata according to CIDOC CRM Digital.

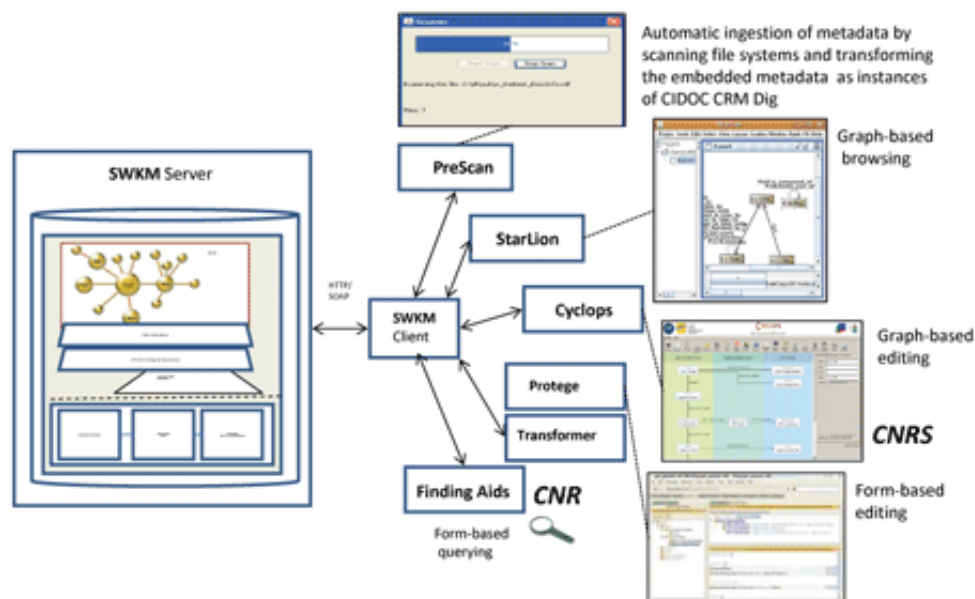


Figure 1: Knowledge Management in Action.

For each of the aforementioned topics, CASPAR has provided guidelines, methodologies and components. Regarding software infrastructure and components, Figure 1 depicts some of the key software components. SWKM (Semantic Web Knowledge Middleware) is middleware for semantic Web data management, GapManager (API and WebApp) offers dependency management services related to intelligibility, PreScan is a tool for automating the extraction, transformation and ingestion of metadata, while a number of other tools for inserting, editing, browsing, searching and visualizing semantic Web data have been designed and developed. The detailed results of this research have been published at DEXA'07, ECDL'07, ISWC'07, ESWC'08, ECAI'08, MEDES'09 and ISWC'09.

Links:CASPAR project: <http://www.casparpreserves.eu/>GapManager: <http://athena.ics.forth.gr:9090/Applications/GapManager/>PreScan: <http://www.ics.forth.gr/prescan>SWKM: <http://athena.ics.forth.gr:9090/SWKM/>StarLion: <http://www.ics.forth.gr/~tzitzik/starlion/>**Please contact:**

Yannis Tzitzikas

FORTH-ICS, Greece

E-mail: tzitzik@ics.forth.gr**ERCIM News 80**

January 2010
Special theme:
Digital Preservation

This issue in [pdf](#)
(64 pages; 15 Mb)



Next issue
January 2016
Next special theme:
Life Science
[Call for the next issue](#)