



Co-funded by the European Commission



Annual Public Report 2008

The main objective of the OKKAM project is to deploy a public global service (called **Entity Name System** or ENS for short) which any application can call to retrieve a uniquely global ID for any type of entity (like people, organizations, locations, events, etc.) named in an electronic information source or document and create customized annotations. The long-term vision of the OKKAM project is that the ENS will enable the Web of Entities, an entity-centric layer on top of the current web where new applications and services can be enabled based on the fact that entities are uniquely identified across any type of format, language, application. The potential of this idea will be demonstrated in three critical areas: enterprise knowledge management, publishing, web search. The OKKAM Project involves 12 partners and will end in June, 2010.

1 Summary of Activities

In the first year of work, the OKKAM consortium has concentrated its efforts in three main directions.

1. Releasing the first running version of the Entity Name System (ENS) and make it available through some APIs. The ENS is complemented with a first suite of OKKAM-empowered tools, namely applications extended to use the ENS for annotating with entity identifiers documents in various formats (MS word documents, email messages, Friend-of-a-Friend profiles, RDF/OWL knowledge bases) and for entity-centric content navigation.
2. Designing the first application prototypes which use OKKAM technologies for fulfilling the identified business needs of the OKKAM application partners (SAP, Elsevier, ANSA).
3. Creating a wide network of scientific and industrial organizations which are potentially interest in using the ENS and related technologies.

The achievements of the first year will be the solid ground for improving the scalability and quality of the ENS infrastructure, developing the application scenarios and increasing the involvement of the scientific and business community in the OKKAM enterprise of enabling the Web of Entities.

2 Important Work Areas

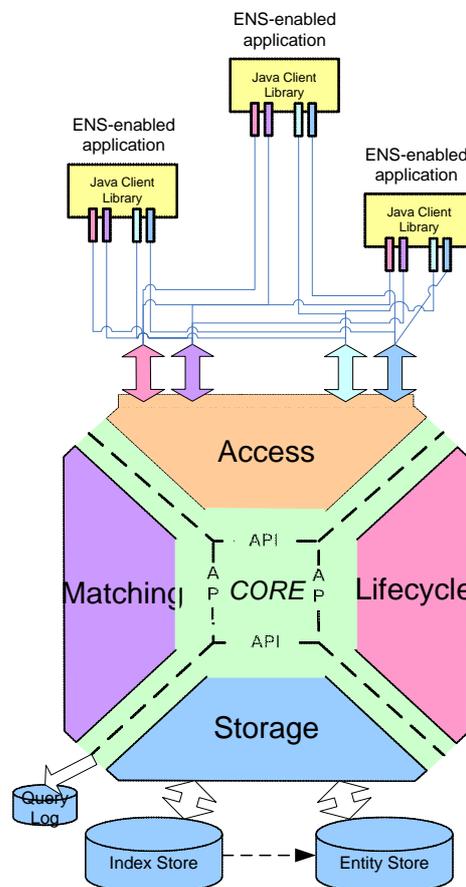
2.1 The Entity Name System, V1

The most important part of R&D work was devoted to deploy the first *integrated* prototype of the ENS. Its deployment required the development of the following core components:

Storage: a scalable repository of entity profiles, in which billions of entities may be assigned an ID and a profile which is used for distinguishing one entity from another. Profiles are mainly name-value pairs, but there is no fixed schema, so they are open, flexible, and type-independent. Part of the information can consist of pointers to other known identifiers for the same entity, and links to authoritative web resources (if available) which describe the entity.

Matching: requests from client applications arrive in the form of a bag of keywords or a collection of name-value pairs (unstructured or semi-structured queries). The matching layer uses this information to send back to the application the ranked a list of the best candidates (if any) found in the ENS. Ideally, this is just one identifier, so we are striving for a very high Top-1 success rate;

Lifecycle Management: it takes care of the evolution



of the repository and of all entity profiles through time. For examples manages the creation of new entities, updates their profiles based on log analysis, provides measures of popularity which can be used to improve ranking of results;

Access Control: it makes sure that only authorized application or users may use the available APIs, filters out mining queries (I.e. queries aimed at retrieving sets of entities based on a common property) and prevents malicious attacks.

All this is made available to applications through a collection of APIs (e.g. for querying the ENS, for creating new entities, etc.), and facilitated through a state-of-the-art, clustered, high-performance architecture that is completely transparent to the user.

2.2 Bootstrapping the ENS and using it

To bootstrap the ENS, a lot of work was devoted to create a significant population of entities in its repository and to enable applications to using it.

Repository population: we have harvested entities (together with an automatically created profile) from some popular public sources like Wikipedia/DBpedia, geonames, UNIProt. In the current production platform of the ENS we have about 1 million entities, but the plan is to reach about 10 million entities by the end of the project.

OKKAM-empowered tools: to make it easy for users to annotate the content they publish with the identifiers coming from the ENS, the idea is to create extensions of popular tools (like MS Word or Outlook) which may enable the user to find the identifier for a named entity and store it with the content in the easiest possible way without getting familiar with new tools. At <http://www.okkam.org/test-tubes> the description of these tools is provided together with a link to the software (where publicly available). So far the deployed OKKAM-empowered tools are: MS Word, Outlook, Firefox, Internet Explorer, Foaf-o-Matic, Protégé, the NeOn ontology editor.

2.3 Application Scenarios

Enterprise Knowledge Management: together with SAP, we are creating an entity-centric application which may support SAP customers to find answers to their problems and experts in the SAP community portals in a much faster and more precise way. The application will identify relevant entities (and their relations) in a user's request and try to match it against past answers to related requests about the same entity (e.g. a specific product) and against experts' profiles. The answer may include material produced outside the SAP community network (e.g. in an informal forum for developers), as external material can also be annotated with OKKAM identifiers.

Authoring and Publishing: together with our partners Elsevier and ANSA, we are creating an authoring environment in which authors (specifically, scientists and journalists) can be supported in creating their content in a faster and better way, and to collect additional information on what they write in a semi-automated way by exploiting the recognition of known entities in their text. Examples can be: proteins in biological papers, past events in news items.

Search: together with DERI Ireland, we are building an entity-centric semantic search engine on top of the Sindice search engine (<http://www.sindice.com>), namely a search engine which may be used to send a query about a given entity (e.g. the European Commission) and will return not just a list of documents, but a structured profile of the requested entity.

3 User Involvement, Promotion and Awareness

During the first year of the project, the incubator Group on RDB2RDF has been joined to push the OKKAM concept and technology into the development of standards and good practices in exposing relational data on the Web in RDF.

The collaboration with the NeOn project produced the NeOn Toolkit plug-in to reuse the global identifier (OkkamID) for individuals in OWL/RDF knowledge bases.

A collaboration with the LarKC Project has been established, to explore the potentials of entity-centric reasoning within OKKAM, and for LarKC.

Contacts were exploited with Nepomuk researchers to evaluate how OKKAM technologies can be used in the Semantic Desktop environment.

The Okkam project organized an *Investor Expert Meeting* held in Brussels, in parallel with the Microsoft SME and Investor Day, and has sponsored the European Conference ESWC2008. It has been involved in the organization of the IRSW2008 workshop, and it will participate at ISWC2009. A full list is given below.

Press coverage mostly referred to the project start, and can be found at <http://www.okkam.org/press-room/press-clips>.

4 Future Work or Exploitation Prospects, as appropriate (ca. 150 words)

In the first year, the goal was to ensure professional communication towards communities of potential users, developers and partners, to foster the creation of OKKAM Communities and to support them, to assess early-on the user, developer and partner models needed to ensure sustainability and development of Project outputs, and to plan the required financing for the continued common research or operational activities.

As a result of the Y1 activities, in March/April 2009 the focus will be on the “going public”. This includes the creation of a *community portal* to foster community involvement, organization of *ad hoc* events, and focused demonstrations. The results of the project will be shown to interested communities, and three applications built on top of the OKKAM infrastructure will be presented (Distributed Infrastructure in V2 and V3 Prototypes). Moreover, the Community Demonstrator “Entity-centric Approach” will be created, together with the Exploitation-oriented demonstrator “OKKAM Results”.

Several OKKAM-related scientific events are in preparation for 2009 which are co-organized by members of the consortium: Semantic Web Applications and Perspectives Workshop (SWAP2009), the 2nd International Conference on Digital Libraries and Semantic Web (ICSD2009), the IJCAI-09 Workshop on Identity and Reference in Web-based Knowledge Representation (IR-KR2009).

5 Further Information

5.1 List of (co-)organized or sponsored Events

PhD Symposium of the 5th European Semantic Web Conference, Tenerife, Spain, June 2nd, 2008.

1st International Workshop on Identity and Reference on the Semantic Web (IRSW2008) hosted by the 5th European Semantic Web Conference, Tenerife, Spain, June 2nd, 2008. (<http://www.okkam.org/IRSW2008/>)

Scalability in Semantic Computing: the European View. Special Session at ICSC2008, the Second IEEE International Conference on Semantic Computing, Santa Clara, CA, USA, August 4-7, 2008. (http://icsc.eecs.uci.edu/special_sessions.html)

The 5th Semantic Web Applications and Perspectives Workshop (SWAP2008), Rome, Italy, December 15-17, 2008. (<http://www.swapconf.it/2008>)

5.2 List of Publicly available Services and Applications

Tough the project is only in its first year, some tools have already been released and made available to anyone who is interested in learning more on what has been done and experimenting with them. Here's a partial list:

- OKKAM Web Search: <http://www.okkam.org/ens/Search.jsp>
Allows search on the ENS repository through a web interface
- Foaf-O-Matic: <http://www.foaf-o-matic.org>
ENS-enabled support to the management of FOAF profiles
- Protège Plugin: <http://www.okkam.org/test-tubes/okkam4p>
Extends the ontology editor Protégé with the possibility of using OKKAM Ids as URIs for ontology instances
- NeON Plugin: <http://www.okkam.org/test-tubes/neon-plugin>
Same as above for the NeOn ontology editor