Interoperability of information systems and the lack of shared semantics is an internationally recognised issue and a grand challenge. Information systems used in the field of higher education in Finland are not an exception. There has been some progress in standardisation of processes and metadata in education technology, but implementing information systems that make use of the standards, best practices and shared semantics is an ongoing struggle.

Typically architecture of higher education information systems is based on traditional databases which have their own data models and lot of harmonization is required to move the data from one information system to another. One proven approach to achieve semantic interoperability between information systems is to build data models and interfaces with re-usable information components [1-3]. Currently a lot of closely related activities are performed by different agencies to define the information architecture and intangible assets. Conceptual modelling of business, services and processes, defining and maintaining controlled terminology and constructing data models for information systems should no longer be seen as separate activities. In order to move from ad-hoc modelling to interoperable data models, a common architecture is needed for semantic interoperability.

In Finland, RAKETTI-project (Information Services as Part of the HEIs’ Structural Reform) [4] defined best practices and guidelines for interoperability during the years 2008-2014. One of the results is the national data model (XDW) [5] that defines concepts and structure for a harmonised database in the field of higher education. The model was designed and implemented to support data collection administered by the Ministry of Education and Culture. It also supports institutions of higher education in implementing data warehouse based on the common data model as data storage for their own operations. Information in the data warehouse can be any functional data of an institution of higher education, ranging from students’ grades to financial indicators. The model is also partly used in the National Data Register of Student Information, Degrees and Credits granted in Finnish institutions of higher education [6].

The XDW model is one of the first common agreements about the data semantics in the field of higher education in Finland. It is, however, already considered to be outdated and the main problem lies in its intended purpose to function as a base for a centralised database. Currently more focus is needed on development of interfaces and data transfers between information systems. As decentralisation of information systems is being preferred, building new data models and harmonizing legacy data requires new approach and tools to build and ensure shared semantics.

In order to move from ad-hoc modelling to semantically interoperable models, a commonly agreed set of terms and shared understanding of their definition are required as foundation. The modelling process [Picture 1] of any information system should start from the conceptual modelling and definition of the relevant concepts to the controlled terminology. The terminology defined in machine readable format, such as SKOS [7], should be requirement for data modelling phase. Of course - in the ready-made world - harmonisation is required in defining and mapping the legacy data models to the controlled terminologies.
Levels of modelling.

Conceptual modeling is used to define the concepts existing in service and process descriptions and information systems. As a method it takes no role in defining logical data structures, such as classes, properties and part-of relations. To be able to build re-usable information components, a semantic interoperability layer is required to link the concepts, international standards and national (logical) data models together [Picture 2]. Currently terminology is often documented in separate wiki pages, documents or spreadsheets, although it is widely advised to define concepts formally to the SKOS vocabularies. Finnish SKOS vocabularies will be published in finto (National Ontology Service) [8] maintained by the Finnish National Library.

Semantic interoperability model is a linked data vocabulary that links SKOS vocabularies and standards, such as ISO MLR [9] or CEN MLO [10] to (logical) data models (Picture 3). The interoperability model defines re-usable information components that are used to create interoperable
Application profiles for specific use cases. Re-usable components are defined as RDF classes and properties and may also be defined as ADMS assets [11]. Logical data models are constructed and documented as application profiles following the guidelines from CEN [12] and DCMI [13]. Following these principles it is also possible to create application profiles directly from the international standards and this approach should be used when creating interoperable data models and interfaces between European institutions of higher education.

New tools and services for developing and maintaining linked data models are needed in order to publish the semantic interoperability model. Standards for Linked Data have been advancing during past years, but tools for creating and publishing Linked Data has been relying on the expertise of specialists and the modeling is usually done by desktop applications. One of the goals of our work is to build prototype for collaborative online tool for creating linked data vocabularies and application profiles. Prototyping is needed to find the requirements and features that enables content specialist to build application profiles and link the models to online Linked Data and SKOS vocabularies.

Picture 3. Linking SKOS and standards to data models and interfaces.

The vision of the semantic interoperability model and re-usable information components is similar to the preceding XML variants CCTS, UBL, NIEM. There is also similarity to the CLARIN CMDI [3] model, which links the profiles to the ISOcat categories. However, the idea of the semantic interoperability model is to connect the (logical) data models to the national terminologies (SKOS) and international standards (CEN and ISO) with unique identifiers (uris) using Linked Data. The idea of the semantic interoperability model is presented in the context of higher education in Finland, but the vision of the linked information components is generic (Picture 4).

Different agencies (e.g. HEIs or libraries) administer SKOS vocabularies, Linked Data models and application profiles. Generic concepts and classifications related to business and operations of agencies are defined in the SKOS vocabularies. The semantic interoperability model defines linked information components which are actually the shared concepts redefined as Classes, Properties and Enumerations. Data models for specific use cases can then be created and led from the semantic interoperability model(s) by defining application profiles. Application profiles created from the semantic interoperability model will be automatically linked to the national terminologies and international standards making it easier to maintain and understand the content of the use case specific data models.

Application profiles are traditionally constructed as human readable documentation, but to make the vision of interoperable future complete, the machine readable version of application profiles is required. The method of defining a machine readable application profile is described in the CEN standard CWA15248 [12] and an updated version of RDF application profile is under development by DCMI Task Group [13]. Formal linkage to the semantic interoperability model and the machine readable format makes it possible to transform the application profile automatically to the standard format. All this ensures the validity of the data by defining constraints and validation rules.
Picture 4. Building the Semantic Interoperability Architecture


[7] http://www.w3.org/2004/02/skos/


[12] CWA15555 - Guidelines and support for building application profiles in e-learning


