

	M	E	T	A	
M	A	T	R	I	X

EMIL – Education Information Markup Language

A metadata model for education information

Metamatrix Development & Consulting AB

	M	E	T	A		METAMATRIX DEVELOPMENT & CONSULTING AB. ORG NR: 556574-1351. URL: WWW.METAMATRIX.SE
M	A	T	R	I	X	SCHEELEGATAN 36, 9TH FLOOR. 112 28 STOCKHOLM. SWEDEN. PHONE: +46 (0)8-33 20 40. FAX: +46 (0)8-33 19 40.

	M	E	T	A	
M	A	T	R	I	X

Table of contents

1	Purpose of this document	3
2	Introduction	3
2.1	From applications to infostructures	3
2.2	A national catalogue for course information	4
2.3	The information hub	5
3	Stakeholders and their needs	6
3.1	Information providers	6
3.2	Information services	7
3.3	Coordinating agencies and organisations	7
3.4	Secondary stakeholders	7
4	Overview	8
5	Design principles	8
5.1	General	8
5.2	Distribution and updates	11
5.3	XML Schema	11
6	End-user services	12
7	Guidelines	14
8	EMIL organization	14

						URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

techniques that would make it possible to handle the “I” regardless of the “T”. The focus shifted from infrastructure to infostructure.

In June 2001 the Ministry of Education and Science commissioned the National Agency for Education to initiate and carry out a strategic program for the development of “soft infrastructures” within the educational system. “Soft infrastructures” were defined as “methods and techniques based on open standards that support and enhance markup, structuring, storing and publishing of information”^[2]. The main goal was to make it possible to communicate and exchange information between heterogeneous information resources. In the commission the Ministry listed both more general activities such as a strategic work to stimulate the use of standards within the educational sector, as well as, if need be, develop new standards, and more specific ones such as developing a national repository for learning objects.

2.2 A national catalogue for course information

Among the more specific activities within the commission was to create a national catalogue containing information about all courses, programmes etc. within the publicly financed educational sector. In the beginning of 2003 a steering committee was constituted with the task to build the national catalogue. The committee comprised representatives from the National Agency for Education, the Swedish National Labour Market Board (AMS) and the Swedish National Agency for School Improvement (MSU).

The project vision was that many different information producers, i.e. universities, university colleges, municipalities (for compulsory and upper secondary school etc.), folk high schools, providers of vocational training, would create and publish, on their own websites, files describing the courses and programmes offered. The files would then be collected by information service providers and used to build various information services (i.e. services aimed at special interest groups or regional services).

At an early stage it was decided that the creation of a national catalogue should be built on two principles: a common information model for describing course information and the creation of a national information service that would collect and distribute the course information files. This service would serve as a base for other services directed at the end-users. At that time there existed already a metadata model which described parts of the Swedish educational sector, developed by the National Agency of Higher Education for their service studera.nu. It was therefore decided that the forthcoming work would build on this model.

The work was thus divided into three parts. The first part was to develop the information model. The model should make it possible to describe, in a structured way, any course or programme within the educational system and to exchange such information among different information resources. The model was given the name EMIL, which stands for Education Information Markup Language. It is expressed as an XML Schema. The second part was to design the national information service, or “information hub”, that would collect and distribute EMIL-files. The third part was to

^[2] <http://mjukis.skolutveckling.se/mjukis/start.jsp?vp=ZA2>

						URL: WWW.METAMATRIX.SE

	M	E	T	A		
M	A	T	R	I	X	

stimulate the development of various end-user services that made use of the information contained in the EMIL-files and distributed by the hub.

2.3 The information hub

There are a number of reasons for making a separation between the information hub and end-user services:

- a hub can cater to the needs of many services thus minimizing the load on the information providers' web sites,
- a hub acts as a cache in the system, if an information provider's web site is unavailable the information is still possible to access for the end user services,
- a hub, especially if run by a national agency, becomes a common and neutral arena to which all information providers can contribute "raw course information".

In the fall of 2004 there exists one national hub, the SUSA-hub, which is run by the National Agency for Education. At present the hub collects fundamental information about the compulsory schools and the upper secondary schools as well as information from the folk high schools. More schools and school forms will be added during the winter and spring of 2005.

Another hub will be developed and put in to operation in the summer of 2005. This hub is being developed for NyA¹, the new admission system for higher education, and will be gathering EMIL-files from all Swedish universities and university colleges. (See chapter 6)

Discussions are also under way in regards to building a hierarchy of regional hubs. These hubs would gather EMIL-files from primarily compulsory schools and upper secondary schools in a certain region. The files would then be passed on to the national hub.

¹ <http://www.nya.vhs.se/>

							URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

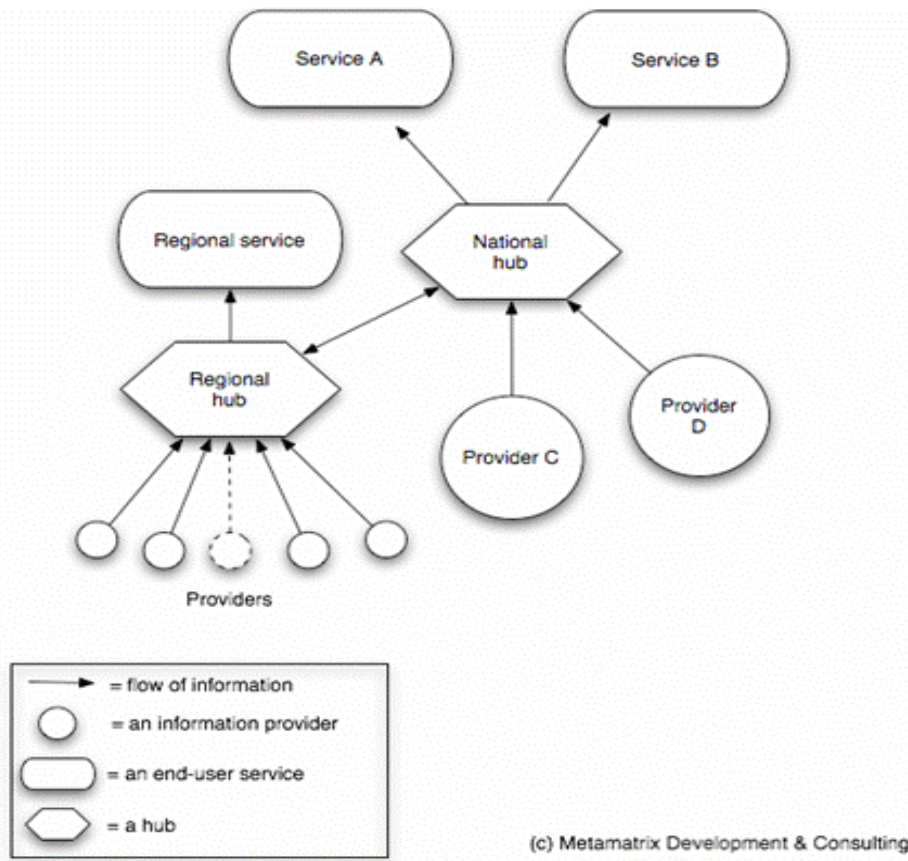


Figure 1 Schematical overview of a hub architecture

3 Stakeholders and their needs

The creation of a national catalogue represents a large scale project, covering the whole publicly financed educational sector with a large set of involved stakeholders with varying interests. This chapter contains a review of the central categories of stakeholders.

It is vital to keep in mind that each stakeholder category will contain a diverse set of members of varying size and with varying resources for IT-based data management, since this influences their needs.

3.1 Information providers

The education providers wish to reach as many potential students or customers as possible with as little effort as possible. They are likely already producing education information as input to local web pages, information material, application systems and other administrative systems and would like to reuse this information if possible.

These stakeholders will benefit from:

						URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

- a common information model - which allow them to export their information once regardless of how many different information providers that may want to use it
- an automated information gathering process - that allows for faster distribution of updated information such as scheduling changes

Stakeholders that have automated their information management may be able to reuse this information as a base for the exported education information.

3.2 Information services

An information service gathers information from one or more providers and provides different types of services based on the information. These stakeholders benefit from:

- machine readable format - that allow them to automate the information gathering
- a common information model - that allow them to use the same implementation for all sources

These stakeholders will have different needs regarding the content of the information model depending on what services they want to provide. The different stakeholders may have conflicting needs regarding things like the amount of detail and structure needed as well as vocabularies for different elements. For example a service that wishes to present available courses within a specific area might want a more extensive and detailed structure of textual presentations.

3.3 Coordinating agencies and organisations

For most types of education there exist agencies or organisations in the public or private sector that coordinate the activities concerning that particular education type. Many of these organisations implement some form of common information service that provides potential students/customers with information about all provided education. The agencies often have a governing role with the purpose of making sure that all providers meet the given requirements on quality and service.

[EMIL-gruppen - finns det regelverk etc. som myndigheter och organisationer behöver stöd för?]

3.4 Secondary stakeholders

While not involved as direct stakeholders in EMIL, potential applicants for scheduled education and potential orderers of on-demand education are central stakeholders for most of the information services based EMIL. This means that their assumed need for information will greatly influence the content of the information model.

						URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

- EducationInfo
- EducationEvent
- EducationProvider

The purpose of working with these as separate identifiable objects is fivefold:

1. Small mobile documents

Aside from being more accessible from a human point of view (easier to read), smaller documents also makes a large amount of education descriptions more manageable for information systems when it comes to distribution or, especially, validation. During an import operation, validating a large document with repetitive components, such as events, may cause unnecessary loss of data, since errors early on in the document are likely to fail the remaining part of the batch.

2. Avoid redundant information

When describing different instances of an education, certain data, such as title and description need not be repeated. This is the essence of the relation between EducationInfo and EducationEvent, as EducationInfo contains information of a more persistent nature for an education. The tricky part in this matter is to determine which information goes where, which has been, and will no doubt continue to be, subject to some debate during the design.

3. Asynchronous distribution

All documents from one information source does not need to be distributed or imported simultaneously, as would be the case with one large document.

4. Changes do not require mass redistribution

New or updated information does only require (re)distribution/import of the documents in question, since they are uniquely identifiable entities of their own.

5. Related objects from different information sources

Globally unique identifiers also allows for related documents to be published from different information sources. The idea was, initially, that EducationProviders might be the most likely document type to be published somewhere else, but recent discussion suggests that there may be other common cases. Some of these depend on the model for relations, see Relations,

5.1.2 Model focus

The EMIL information model started out as a very data-centric design, whose whole purpose was describing searchable data that was to be delivered to an information system. As the model has developed, in close connection to systems being implemented,

						URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

it has adapted to meet new needs. Notably, more descriptive elements have been added that are for displaying rather than searching.

5.1.3 Relations

An ongoing discussion has been the concept of relations between documents. In the current proposal, the EducationEvent has a relation to one EducationInfo and one EducationProvider. This provides the most basic function, to create a connection between three documents that together describes a complete instance of an education, at least, complete for less complex forms of an education. One issue that is already on the agenda is the possibility to relate an event to more than one provider, as this is the case for some academic courses.

Another aspect of this subject is the ability to express hierarchical relations. For instance, a local variant of an education may be described with a relation to a general education description. This has however not yet been incorporated in the design. The current design allows an event to have relations to other events. As of yet this is a construction used for representing packages of academic courses. The relation design is thus an area for further discussion.

5.1.4 Vocabularies

As in most cases with metadata markup, some textual information elements in EMIL are constrained to a list of values that are valid. These information elements often indicate a “type” for something, for example “type of activity”, “type of level” or “course type”. A decision was made early on to exclude all vocabularies that were judged too large or frequently modified. The “subject” vocabulary for educations meets both of these criteria, as it tends to be a rather extensive list that changes regularly. Instead, these vocabularies are managed by stakeholders that have information systems with specific demands.

The overall use of vocabularies is an open issue. Some vocabularies in EMIL are still in Swedish, mainly because they are less than simple to translate into English. Others are cause for disputes as they vary greatly between different types or levels of education. International standards are of interest in the continuing work.

5.1.5 Languages

A few descriptive elements are language aware, that is, it is possible to include textual descriptions in more than one language. The specific language is indicated by an attribute which contains an ISO 639 code (the xs:language-type).

5.1.6 Extensions

No model is ever complete for all users. Therefore, being able to extend the model and include user-specific data is a necessity. Each of the three document models accepts externally defined elements, as many as desired, currently all in one dedicated area.

						URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

5.2 Distribution and updates

EMIL documents are meant to be distributed in a network as described above (information hub). In this kind of system there are a few pieces of information that are necessary for the overall handling of validation and updates of documents. First and foremost, we need a way to identify each document uniquely in a global context. Otherwise we would undoubtedly get conflicts and risk incorrect information. Secondly, we need to be able to discern whether an incoming document has been changed since the last time we fetched it. It is also important to know when information is obsolete, i.e. when an EMIL document expires. These three parts are defined as attributes in a document:

- id - a unique identifier for this document
- lastEdited - date and time when this document was updated
- expires - date when the information in this document is no longer valid

Additionally, there is one attribute that is always present in an EMIL document, namely

- version - fixed version of the XML Schema

5.3 XML Schema

Each document type is defined in an XML Schema (W3C). The schema design is basically a straight forward venetian blinds, with the only allowed root elements as globally defined elements.

5.3.1 Naming conventions

Elements and attributes are named as believed best suited for announcing the purpose of the content. On a more technical note, lower camel case is used for both element and attribute name.

5.3.2 Namespaces

The namespaces name for each document type is a URI that consists of:

owner domain + EMIL document type + version

Example: <http://skolverket.se/education/event/0.3.2>

5.3.3 Versioning

Since EMIL is still a work-in-progress, new versions has as of yet not been backwards or forwards compatible. That would have been too limiting when modelling. Future development will aim for backwards compatibility for minor version upgrades and for forwards compatibility between minor versions in cases where new (added) elements are not used. New major versions will not guarantee compatibility. The version number is currently included in a fixed attribute on each document, and in the namespace for the document type. By including the version number in the namespace

						URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

AMS

Since 1998 the Swedish National Labour Market Board (AMS) has an information service on its website^[4], allowing the users to search for Swedish public courses and educations (from upper secondary school courses to university education, except from free-standing courses within universities). For the time being there are approximately 4000 different educations available in the database. The search can be carried out with a choice of education area and/or -direction, type of school, county/municipality, school term and distance/flexible tuition. The different educations are today updated manually, although the agency's objective is to, with the use of XML, make the gathering process more efficient and thereby keep the information more topical. The goal is that during the autumn of 2005 the technical environment should be adapted for the gathering and presentation of education information on the website.

Other national services

Other end-user services that will build on the information contained in EMIL-files are the course catalogues at the Swedish Net University Agency and the Swedish Institute. Both these catalogues present a subset of all available courses and programmes; in turn, all academic courses and programmes, all academic courses and programmes offered as distance education and, all programmes open for students from outside Sweden. These catalogues will get their (filtered) information from one of the national hubs, most likely the NyA-hub. It should be noted that the agencies mentioned here already have functioning catalogues today. These catalogues use XHTML-files and an expanded set of Dublin Core metadata. The course files are gathered from the universities and university colleges only by the National Agency for Higher Education, which then relays the information to the two other agencies. This system is indeed the fore runner and source of inspiration for the more ambitious "EMIL, hub and end-user services"-concept.

Whereas the previous services have presented subsets of all available courses and programmes based on educational level or form, there are also services that present other subsets. One is the INSET-service for in-service training of teachers. This service is managed by the National Union of Teachers in Sweden. Here the subset of presented courses and programmes is not so much defined by form or level on as what a teacher may need for in-service training (e.g. programmes of any kind will not be shown in this service). This service is in a test phase using the old DC-model but will transfer to the EMIL-model during 2005.

Regional services

Another type of services is regional services built on the regional hubs. Such services will provide a complete picture of the courses and programmes, regardless of level, offered in a region. The service can of course also gather courses of interest for students

^[4] <http://utbildningar.ams.se/main/frames.asp>

							URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

in the region but not offered in the region or offered as distance education from another regional or a national hub.

7 Guidelines

Since EMIL is an extensive model it therefore has to be general in its design at the same time as it has to address all aspects of describing education from when to how. The idea is that the services that accept EMIL should make demands what is needed for them. There are therefore few obligatory elements, every stakeholder can decide what their obligatory elements should be.

Every instance can decide if there should be specific vocabularies to follow. One example is the subject scheme for higher education. There are specific needs that have to be addressed in that area that is not applicable for general education. This can easily be solved with specific vocabularies. (See 5.1.3)

In some instances specific extensions are needed. These are often used when there is an immediate need for one area and the time does not allow for a new version of the scheme. (See 5.1.5 and 5.3.4) Members of the EMIL group can decide at a later date to let the extension to be introduced into the next version of the scheme.

There is a document that describes the implementation of the EMIL scheme in Swedish that has of yet not been translated into English. This document describes in detail how to set up a hub of your own or how to report that you have files to share with the community.

8 EMIL organization

The work is lead by a steering group with a represent ant from each participating Agency/organization. The steering group decides on the work plan and new versions of the information model. The group decides upon a chairman who leads and coordinates the work. Decisions can be taken both at meetings and after e-mail correspondence and always as consensus decisions. In the case of e-mail decisions every member has to give their vote.

To help the steering group there is the reference group, where anybody who wants to participate is welcome to do so. Their first and foremost mission is to be a referral body for any change in the information model. Any change of a substantial nature should be sent for referral to the reference group.

The development of the information model will be co-ordinated by one organization as the orderer, in case of external help is needed. Ordering of such work should be notified to the chairman and any demand specification should be sent to all other members for referral.

Decision for changes should be taken in the steering group after taking into account every members point of view. As far as it is possible this work is to be conducted in the open.

						URL: WWW.METAMATRIX.SE

	M	E	T	A	
M	A	T	R	I	X

Aim for the work is as follows:

- Develop an information model (version 1.0) that describes education information education event and education provider
- Develop a XML scheme (version 1.0) that expresses the model
- Develop information and presentation material for the model and the schema
- To continuously inform all stakeholders of the model and its applications
- To develop rules and guidelines on how the model should evolve, how the work should be financed, how to decide upon the versions of the model and how to divide the work between the actors
- To continuously develop both the information model and the scheme
- To work towards coordination between the information model and similar initiatives foremost in the European community
- To continuously develop cooperation between Agencies/organizations involved in EMIL

There is a Swedish version of the work plan on:

<http://mjukis.skolutveckling.se/mjukis/start.jsp?vp=ZA21ZDL>

						URL: WWW.METAMATRIX.SE