WP2 Note on ECHO's Digital Open Resource Area DORA

Peter Wittenburg
December 10, 2003

1. DORA Design Principles

DORA is the portal that offers discovery services for various resources that were and are created by major European initiatives, in particular by the ECHO initiative. The **ECHO** initiative is gathering resources in the five different disciplines Linguistics, History of Art, History of Science, Ethnology and Philosophy.

Under the header of **Linguistics** resources from a couple of other initiatives will be made available as well:

- the **INTERA** project that has as goal to create an integrated domain of language resources;
- the **DOBES** project documenting endangered languages all over the world;
- the **MPI and the Lund University** language resources.

While the linguistic part in ECHO focuses on minority languages such as Sign Language and linguistic objects with a heritage aspect, INTERA is focusing on major languages and combining language resource centers in Europe and DOBES is focusing on languages (in particular non-European) that probably will become extinct in a few years time. In combining these initiatives and the MPI for Psycholinguistics as well DORA will offer access to a large set and therefore forming a critical mass.

Under the header of **Ethnology** also various resources will be made available: the NECEP society database, the collection of the DOGON project and the large collection of the Dutch Ethnology Museum (RMV). Other resources may be integrated as well at a later stadium.

In the area of **History of Arts** three databases will be added: Fotothek, Lineamenta and ancient maps of Rome. All are housed at the Bibl. Herziana.

In the area of **History of Science** a number of collections will be part of the DORA domain. IMSS Florence will contribute with its large collection and institutions such as U Bern, MPI for History of Science and perhaps others will contribute as well.

In the area of **Philosophy** the collection of texts from the ECHO partner will be integrated.

DORA offers various access methods primarily to the metadata descriptions as a simple and easy navigation space. Hits will allow the users to access the resources themselves, given that they have the proper access rights. The metadata descriptions are openly accessible. The access to the resources that can be text, images, movies, sounds and 3D objects may be restricted. Various views and access mechanisms will be available to meet the requirements of the different user groups.

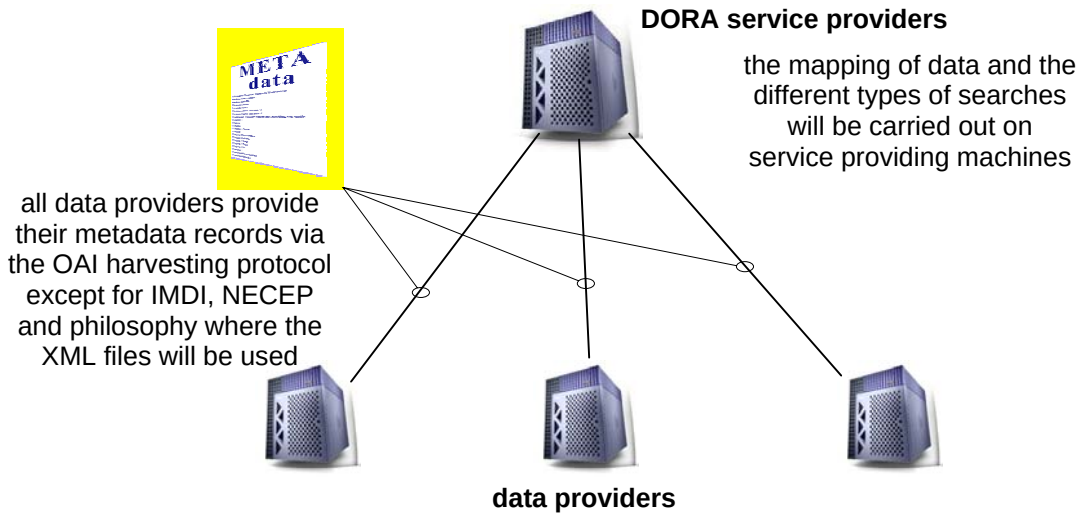
The language resource domain within DORA is mainly using the IMDI metadata standard, although this is not necessary. Therefore, the IMDI domain is a large sub-domain in DORA. For many other holdings different metadata sets are used, i.e. to create a unified umbrella various mappings have to be carried out. This is described later in this document.

At first instance Lund U and the MPI Nijmegen will maintain DORA. However, others can set up a similar portal since the sources will be made openly available.

1.1 Topology

The DORA service is a central one, i.e. all metadata will be harvested at a central server and stored optimally for fast access. This implies that the central server will only have copies of data, the original copies will stay at the original institutions where they also may be subject of changes and extensions. With each partner a procedure will be discussed that will allow us to harvest the metadata records. The DORA service is not a service that extends to the resources themselves, i.e. the metadata may have references to the digital objects they describe such as images, texts, sound files or movies, but these resources stay at the institutions. If a certain institution does not have sufficient resources to house videos ECHO could act as an umbrella to also house the resources at a central server[1].

Summarizing we can conclude that in the DORA metadata scenario all institutions act as data providers, i.e. they offer their metadata records for being harvested by the DORA service providers. Different protocols will be necessary to harvest the data. Different types of records will be offered by the different institutions.



1.2 User Interface Aspects

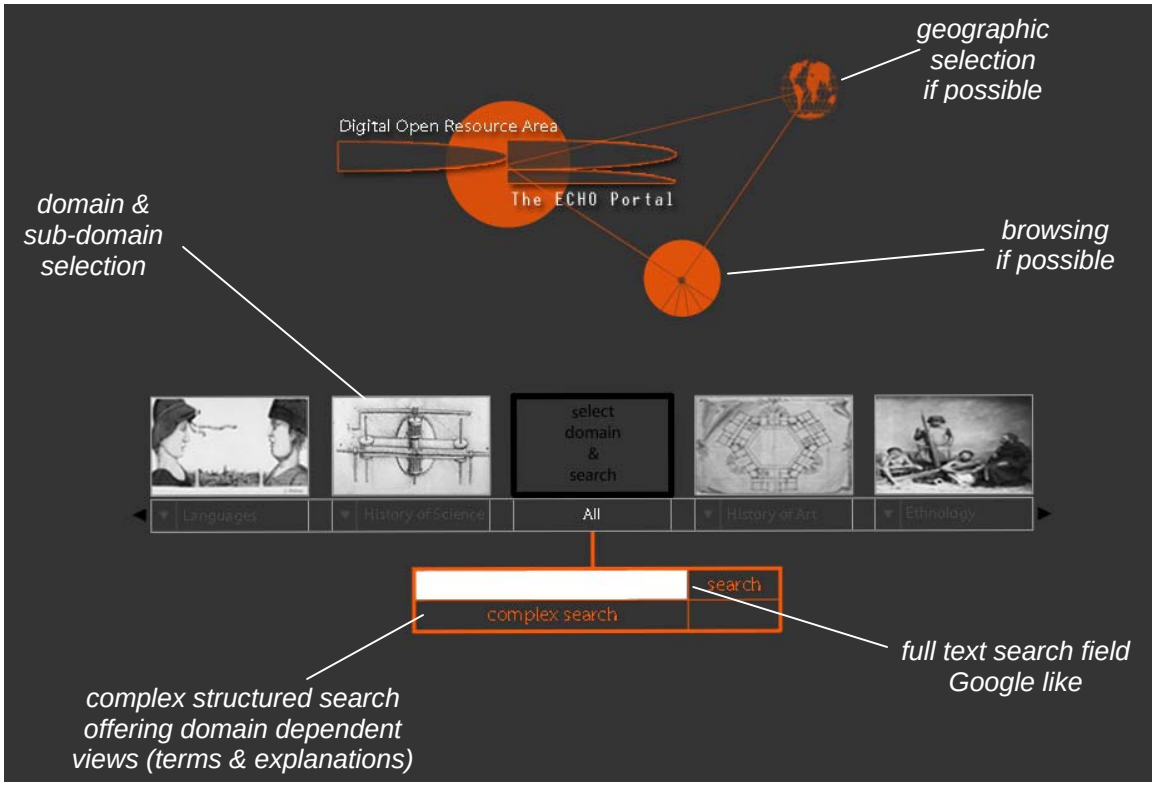
First we want to list a number of requirements for the user interface:

- it has to support the normal working environments such as web browsers (first a limited set of browsers will be supported)

[1] Under certain circumstances the MPI for Psycholinguistics could house resources.

- it has to be simple and robust
- it has to look professional for the normal web user
- it has to offer simple Google like search on metadata as the first choice[2]
- users can select the domain they want to search in - the default domain is "all"
    - a preference file has to support that different users have different defaults (question where to store this: on server or as bookmarks, ...)
- users can select a certain view (domain specific vocabulary) to specify their queries
- the opening page has to be attractive, i.e. the layout has to be designed carefully
- all pages must use one underlying style
- the opening page has to
    - allow to jump to geographic browsing (no idea yet whether we can include other resources than from languages and ethnology)
    - allow to jump to IMDI type tree browsing
    - allow to go to the specific search engines provided by the disciplines such as the full IMDI infrastructure
- the opening page should contain all relevant links (ECHO, IMDI, MPI, DOBES, ELRA, Lund, INTERA, ...)
- it has to be checked in how far we want to extend to DC/OLAC repositories, i.e. in how far we want to harvest other sites
- the DORA service should allow OAI (DC) service providers to harvest its holding
- the first version must be ready as soon as possible, i.e. when components are ready they should be made visible

DORA Main Page
(test page is available under: corpus1.mpi.nl/ds/dora_demo2; please, note that it is under construction)



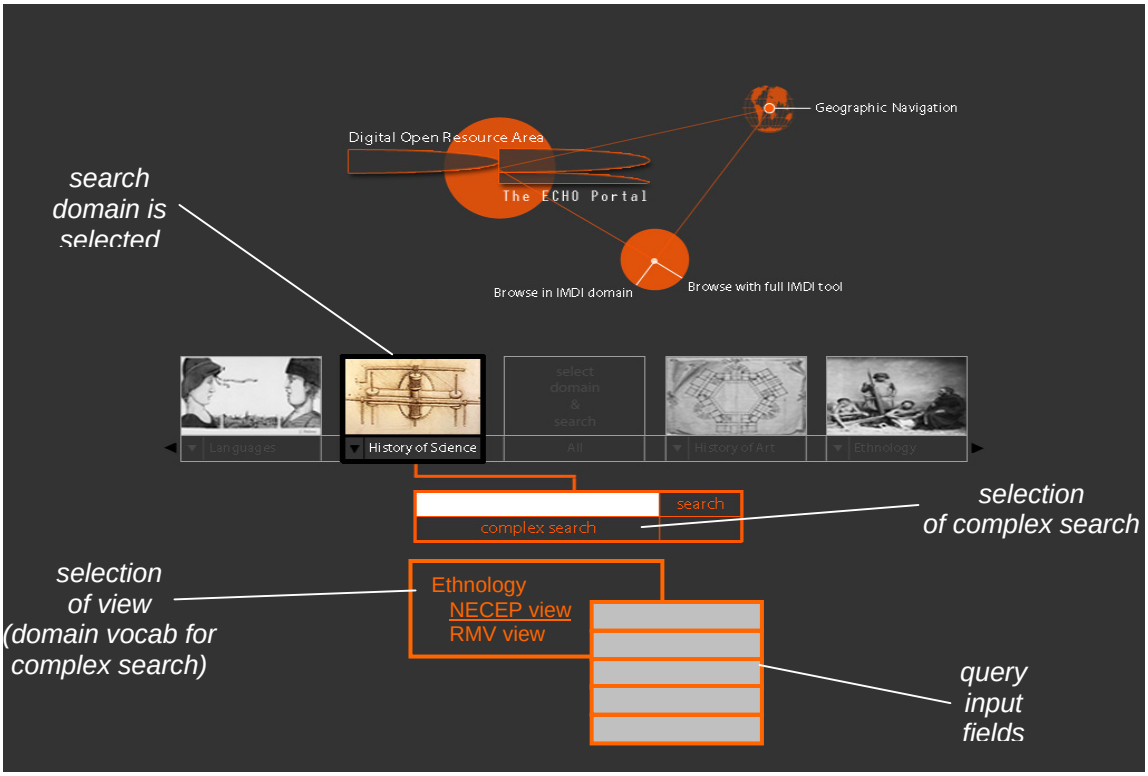
This figure[3] indicates the major elements of the DORA user interface. It will support simple search, complex structured search, selection of domains and where possible geographical and

[2] In a second version a lexicon could be displayed to help people to find suitable terms while indicating the domain from which they are taken.
[3] Yet an appropriate symbol representing philosophy is missing.

hierarchical browsing. In this version we miss an indication of the possibility to extend the simple search on metadata (keyword type), annotations (general type of metadata) and relations.

For all forms of searches (simple and complex) the terms used in the descriptions will be indicated in a separate window. This will facilitate searching since it will inform the user about what is existing and it will minimize typing errors. It has to be worked out what the best way is to offer the wordlist in a structured way since they can become very long.

Complex Search Page
When the user selects Complex Search the following page will show up:



Still the user can select the domain and sub-domain he/she wants to search in and whether he/she wants to search on metadata, annotations and/or relations. When a special view is selected a suitable vocabulary will be shown which the user may be more familiar with. The offered fields can be used to enter strings to form the structured query. In general we will use a subset of elements from the different domains. Candidates are such elements that can be mapped to other domains. If users want to do more specific searches using elements that cannot be mapped they will be able to go to the specific search engines.

One of the detailed views is the DC view and it will offer the well-known 15 DC elements.

Browsing Page
Currently, we see two domains where browsing in metadata domains is an issue. IMDI uses this concept for language resources and the Alcatraz environment seems to support browsing according to some thesaurus. Where possible we will support browsing in such metadata domains.

An interaction should be supported in so far that any browsing is used as a specification of a sub-domain for simple search as well. If a user has selected some node by browsing it should therefore be possible to do simple search and use the node as a selection criterion to narrow down the search space.

Since date information is used by many metadata sets it has to be checked in how far it is possible to generate a browsable tree that orders resources according to their date.

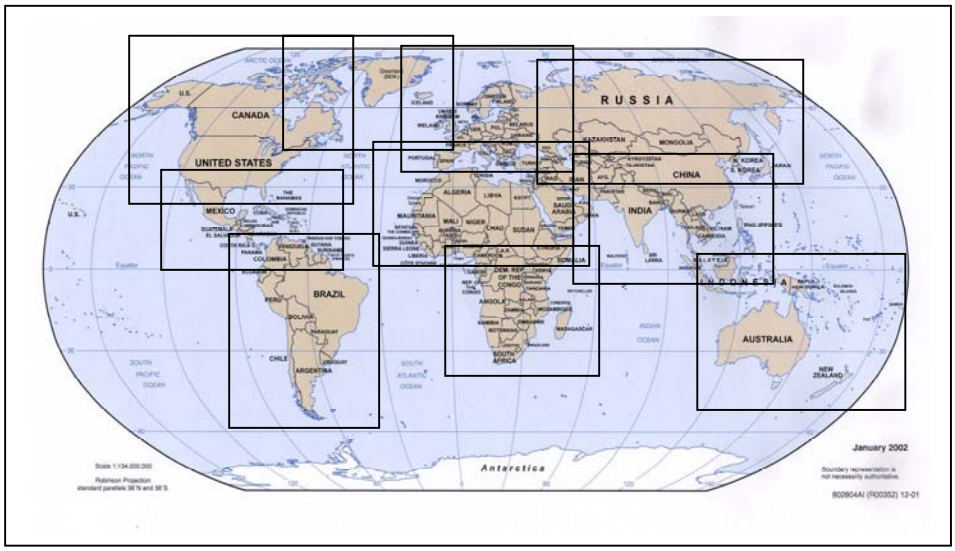
Geographic Browsing Page

One very popular form of browsing is to use geographical information. Since many metadata sets are using geographic indicators such as continent, country, region and place it may be possible to add this type of information to geographic maps such that people can make selections based on these maps.

DORA has to differentiate the different usages of the geographical information, i.e. the place of origin is not the same as the place where an object is located. In general one would use the place of origin within the DORA framework. This has to be analyzed in more detail.

Again here it is important to allow selection criteria, i.e. to only show information for the selected domains and sub-domains. In many cases it is a problem to associate a document with geographical maps. A society will live within a region, but drawing regions can easily cause

political problems. Therefore, DORA will associate information with useful points on the maps although this is not as optimal in many respects.



The world map can be broken up into a number of sub-pages at two or three levels. A possible second layer is indicated in the figure above. That should be sufficient to mark all points with sufficient detail. There may be some detail maps as for the History of Arts where most resources point to places in Italy. When selecting a point by clicking all resources are shown as hits such that people can view or listen them.

1.3 Selection & Searching Modes

Here we want to summarize the searching modes again.

- **Domain Selection**. The user can select the domains he wants to operate in and that has to affect the search and selection modes except the geographic one. We will offer domains and sub-domains for selection.
- **Resource-Type Selection**. The user can select to operate on metadata, annotations and/or relations in the simple search modus.
- **Simple search** offers Google like facilities and at first instance the user does not get any help. At a later stage one could think of a lexicon of all possible terms. This simple search operates on an index that contains all metadata values that occur in the participating domains. This includes in particular the descriptions since for example in ethnology especially the descriptions contain the useful material. In doing so ss ignores all structure of the metadata sets and therefore looses the high precision of structured search.
- **Complex Search** offers a few major categories of each domain with a domain specific naming. In particular those categories that can be mapped between the disciplines should be mentioned. It has yet to be defined which categories will be made available. Of course, in this mode the controlled vocabularies should be available to guide the users.
- **Browsing** can be chosen to navigate in browsable domains such as the IMDI world with normal web browsers making use of on the fly created html. The possibility of automatically creating a historical browsing tree will be investigated.
- **Geographic Selection** can be chosen by clicking on the world map. The only possibility is to click on marked spots that will result in a list of all sessions belonging to this spot and display them. It has to be checked in how far this can be improved by linking to a node in browsable trees. So - clicking on a spot in the map will execute a complex search with the location and or item information (this has to be carefully checked).
- **Domain-Specific Search.** The user has the possibility to go to the domain specific search that will offer all fields for that particular domain or sub-domain.

Use of Mappings
Since DORA will combine different domains, terminologies have to be mapped while searching. The detailed mappings have to be worked out. The mappings will be used when performing a complex search. In simple search any term can be entered and the program does not know which view the person takes. So term mapping does not make sense for simple search.

In complex search a user takes a view. This activates a number of mapping tables from the chosen user views to the other domains. The mappings will extend and modify the search query for the other domains.

1.4 Domains und Sub-Domains

DORA knows a number of domains and sub-domains. They can be changeable in a domain configuration file.

The Domains and Sub-Domains are:

- Languages
  - ECHO
  - IMDI Domain
  - INTERA
  - DOBES
  - MPI Nijmegen
  - Lund
- Ethnology
  - NECEP Paris
  - DOGON Leiden
  - RMV Leiden
- History of Arts
  - Lineamenta
  - Fotothek
  - Ancient Maps of Rome
- History of Science
  - IMSS Florence
  - Collections from Bern and Berlin
- Philosophy
  - Philosophy Paris

The domain-configuration file has to include addresses that can be used for harvesting purposes as well. This configuration file can be used to generate the entries and menus. An indication is given below. The details have to be worked out.

domain-name	sub-domain-name	protocol	address	web-site	cv addresses

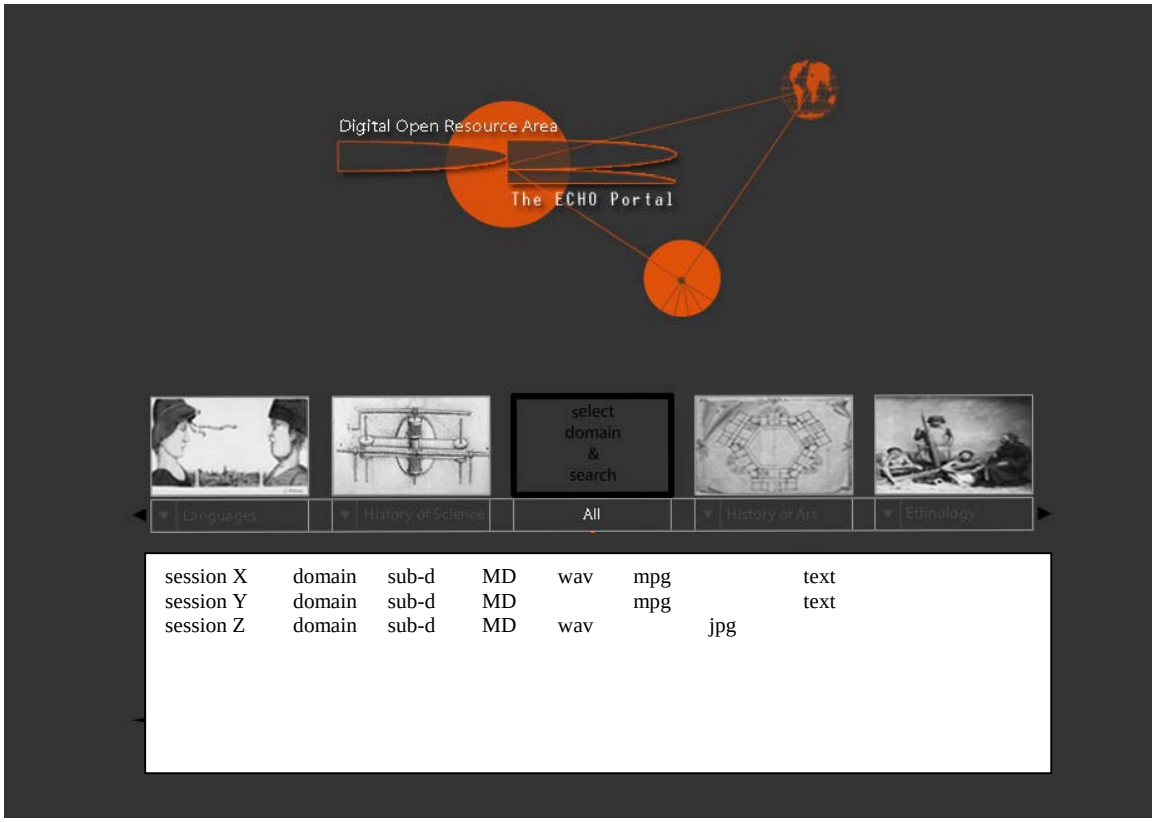
1.5 Hitlist

All hits as search results have to be shown in a unique way offering the DORA style and a number of choices. The web site should immediately allow to continue searching etc, i.e. the actual selection and navigation mode should be shown again. Here we can learn from Google to optimize ergonomics.

From the hit list it should be possible to

- view the metadata record and from there jump to other sources such as info files or articles (references)
- view and listen to the resources
- invoke other shells that allow to go on with navigating and visualization (this has to be discussed in detail how it can be done)[4]

In the case that it is not possible to directly refer to the resources a suitable shell from the participating sites has to be invoked with the correct arguments. For streaming audio/video a communication with a streaming server has to be realized.



The layout for the hit-list page is only indicated schematically. The presentation as a simple list is not at all optimal, since people want to exploit results in a more suitable form. But in the first version nothing special will be done. Google-like designs should be considered.

At first instance there is no rating involved. Due to the involvement of different domains we first have to get experience with result lists. Different domains may require different criteria for determining the relevance of a document.

Possible criteria could be:

- hit comes from structured vs. non-structured information
- weak mappings are indicated and drop the rating
- spelling differences between terms
- frequency of terms found in a metadata record and in associated documents

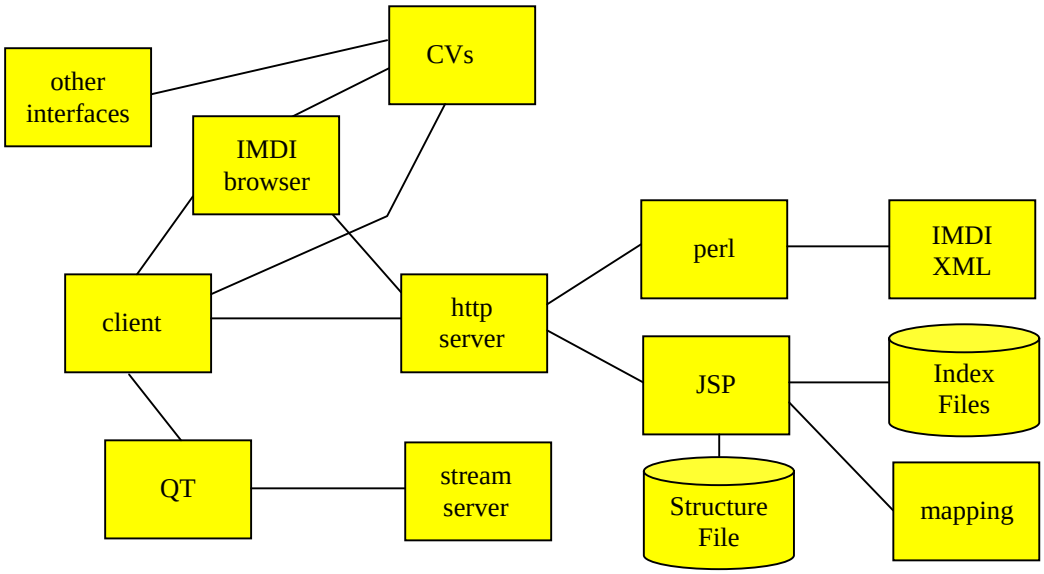
This has to be sorted out in a later phase.

1.6 Implementation Issues

At the client side normal html and JavaScript is used. For streaming services the QT client has to be invoked (QT has to receive the right parameters to be able to request the execution of a

[4] Users may want to go from a hit for example about a DOGON building directly to images or to the guided DOGON tour that is available at a web-site.

certain file) and for example for full IMDI requests the IMDI browser can be used. It has to be checked in how far controlled vocabularies have to be used to support structured search or whether it is better to offer the actual terms used. At the server side Perl/XSLT scripts will be used to generate the html information that is extracted for example from the IMDI and other XML files.



JavaServerPages will be used to solve all other aspects at the server side. It will access index files to quickly generate results in the two searching modes. It has to be sorted out whether the full text search will need a different kind of index structure than that one that is used for the structured search. JSP need the mapping files for cross-discipline activities.

JSP need the IMDI structure file to support the restricted search that was described on the browsing page. When someone is browsing for example in the IMDI domain a selected node could be the start for an additional search, i.e. this requires that the selection made is known to the JSP. To restrict the search JSP have to know which sessions belong to that node.

Perhaps controlled vocabularies have to be supported in the second phase. In the configuration file all CVs used have to be specified by its address and the category it is associated with.

1.7 Harvesting Comments

With respect to the harvesting some general comments should be made for clarification:

- Only data from known sites will be harvested, i.e. data on local notebooks or so are not considered.
- The amount of searchable data can become fairly large, in particular if we integrate annotations and relations.
- We assume that the repository content will change, i.e. harvesting should be carried out at regular intervals. This has to be discussed in more detail with the partners depending on the experiences.
- The MD schemas may change. Special attention has to be drawn to such occasions.
- Keyword-value pairs as possible in IMDI will be treated as descriptions at first instance.
- Those who chose to be harvested via the OAI harvesting protocol have to register as OAI data providers. MPI for Psycholinguistics can offer help.

2. Metadata Mapping

WP2 has to realize an infrastructure for joint searching and where possible browsing covering all disciplines in ECHO: history of arts, history of science, ethnology, linguistics and philosophy. The metadata sets applied in the different fields are different in many ways such that mapping is required. Further, the interface has to be offered in several languages such that dedications of all terms to these languages are required. We also have to accept that at this moment the used element names are not yet defined in open repositories according to international standards such as for example ISO 11179. We lack appropriate and accepted tools and repository structures.

Therefore this note suggests preliminary structures for open repositories (available at the WP2 site) that contain element definitions, translations to some languages and relations between the elements. The information has to be such that it can be easily transformed into future frameworks. In this document version we will not yet translate the schemas into RDF, but first describe the structures in XML. The RDF formulations will be added later. What we will do is to describe the immediate requirements resulting from establishing a common search infrastructure.

2.1 Introduction

We are faced with several domain and sub-domain ontologies that all use their own definitions of elements (terms), i.e. there is nothing as a common ontology. Therefore, within ECHO we have to develop a framework that allows the mapping between the different metadata sets.

First, we would like to briefly characterize the metadata sets of the participating domains/sub-domains.

domain = languages
- all metadata is filled in according to the IMDI standard; so sub-domains are included just as other linked IMDI repositories;

sub-domain = -
- all contributors share the same element semantics
- the metadata set includes a rich description that describes the project, the researchers, the formal nature of the resources and their contents; it contains about 40 elements and points to the raw and derived resources
- the metadata set was designed to manage and discover resources in large distributed scenario
- the number of metadata records is currently larger than 20.000; due to ongoing work this number is continuously increasing;
- for the metadata details see www.mpi.nl/IMDI

domain = ethnology
sub-domain = NECEP (database of societies)
- with the help of an exhaustive set of elements (about 150) researchers are describing societies; in addition prose texts elaborate on certain aspects of societies and explain how to interpret the chosen values; where possible additional media resources illustrate aspects;
- the metadata set was designed to describe societies in great detail and also to easily find information on societies;
- the database is in its beginning phase, i.e. there are only a few records and the expectation is to have about 10 controlled ones at the end of the ECHO project;
- for the metadata details see ???

domain = ethnology
sub-domain = Dutch Ethnology Museum (RMV)
- RMV has a huge collection of ethnological objects (>250.000) of which only a few are available in digital form and described by metadata (> 3500); every year the digital collection increases in size by about 3500 objects;
- for budget reasons only 12 elements are used to describe the objects;
- metadata is used to easily discover objects in the digital archive;
- for the metadata details see appendix A

domain = history of arts
sub-domain = fotothek database (Bibliotheka Herziana)

The Fotothek is a large collection of partly related digital images (6.000 images, 100.000 descriptions); all images are described by metadata that are created according to the MIDAS standard that uses the IconClass thesaurus to encode the content;
the MIDAS standard is an exhaustive set that has elements to describe the creator, the involved archives, the content ??; it also encodes hierarchical relationships;
metadata is used for management and discovery purposes;
for the metadata details see appendix D

domain = history of arts
sub-domain = lineamenta database

The lineamenta database is a new database, its new integrated design was developed to include all sorts of information; survey type of metadata is included in different tables; internally they use a rich metadata set, but only comparatively few fields will be exported to fit with the metadata scheme introduced by history of science (see below); in total there are 500.000 drawings, but the project assumes that at the end of the ECHO project about 300 drawings will be described; internally

domain = history of arts
sub-domain = ancient maps of Rome database

The maps of Rome is currently a small database of about 200 maps described with the help of metadata, the detailed set has to be investigated in more detail, first data was provided.

domain = history of science
sub-domain = Berlin/Bern

The metadata set is a new one and contains about 30 elements; it is possible to add another 15 elements taken from Dublin Core;
most of the metadata elements are used for administrational purposes, i.e. only few can be used for resource discovery, in particular in cross-discipline environments;
for the metadata details see appendix B

domain = history of science
sub-domain = IMSS Florence

IMSS has a large collection of instruments, documents and artistic objects all being catalogued; recently a new metadata set has been worked out that uses the Dublin Core field as the core and has for each document type a couple of extra fields, therefore the total amount of fields is about 40 and the set is flat, IMSS just started to fill in these templates to describe their holding

domain = philosophy

The philosophy domain does not have sub-domains; the philosophy group from Paris is working on a fully-linked rich dictionary that translates “terms” into different languages; there will limited set of lexical entries (terms) at the end of the ECHO project; typical metadata fields are used to describe the lexical entries; a precise set is being determined currently – it will be extracted from the texts

2.2 Metadata Elements for DORA[5]

DORA offers a number of ways for searching: full-text searching on all metadata elements (and even beyond keyword type metadata), structured search offering selected elements and geographical search where possible. For people with detailed queries the portal will link through to the specialized sites.

All ways of searching are based on metadata (and partly on annotation) harvesting. The DORA service provider applies two methods of harvesting as described in chapter 1.1. The DORA

[5] DORA = the ECHO portal called Digital Open Resource Area

service will harvest complete records such as they are offered by the data providers. Filtering and indexing as necessary for the different search options will be done by the DORA service.

It has to be checked in a second phase how the annotations and relations will be harvested. At first instance they don't fit with the OAI model, since the required Dublin Core set cannot be provided – so registration as OAI data provider is not possible. If data is openly available and in XML format the most easy way would be to read the XML files.

2.2.1 Full-text Search

For full-text search we will include all fields of the different metadata sets and optionally annotations and relations. We assume that those fields that don't bear meaningful information to be queried such as addresses, references/links, contact names etc will not decrease the precision and recall significantly.

The DORA service provider will harvest[6] all metadata information that will be offered by the data providers and for full-text search create joint indexes. These will be created such that we can trace back from which domain and sub-domain the hits were taken.

For full-text search there are no different views, i.e. no specialized domain-specific vocabulary. The consequence is that full-text search does not support semantic mapping at first instance. The search should offer a wordlist, however, that shows the user the possibilities when typing his query. This feature can be used as well for checking typo errors and for easy completion.

2.2.2 Structured Search

To support structured search we have to be selective and only support those elements that can be mapped between the different domains and sub-domains. We can expect that the user who wants to search for domain-specific details will always want to use domain-specific interfaces.

For inputting and executing queries two options have to be available:

- The user must be able to select the domains and sub-domains the search should include.
- The user must be able to select a view (terminology) to input his query. Since there are even large differences between the terminologies used by the sub-communities, the user must be able to select a sub-community view.

In addition to the domain/sub-domain views we will add the Dublin Core view that will offer the Dublin Core vocabulary. The table below gives a first idea of which field will be used from the different domains/sub-domains and how they can be mapped. Since there are so many differences between the domains we started with dualistic mapping schemes between two sets and from there derive mappings for each view. In the table we use the mapping from Dublin Core to the other domains serves as a basis for explanation. We have to develop such mapping schemes from every view since yet we cannot identify a common base such as is used in WordNet that uses a common list of concepts.

At first instance we will exclude the unmarked fields (white) from the view since they don't seem to offer very promising results.

From this exemplary table it is obvious that the semantic mapping of the metadata elements is not at all trivial. The decisions made can lead to misleading results and wrong conclusions. Therefore, it is necessary to allow people to use other mapping schemes. This would mean that it must be possible to either make it easy to set up a new service provider or to influence the logic machine by pointing to different ontologies.

As an example for the problems we will discuss in the following paragraphs three cases are discussed:

[6] Harvesting will be done by requesting XML files using HTTP or by applying the OAI MH protocol. The details are described in other WP2 documents.

- the more simple one of “geographic location”
- the slightly more difficult one of “languages”
- the more difficult one to map content

DC	Ethnology		History of Arts		History of Science		Languages
	NECEP	RMV	Fotothek	Lineamenta	Berlin	IMSS	IMDI
Title		object name	object title	title	m.title	title	title
Creator			name artist	person	creator m.author	creator	participant
Subject		categorization	title of building prim icono sec icono	object keywords	keywords	subject	content language
Description							
Publisher							
Contributor			name artist	person	m.author	contributor	participant
Date		date	date period	date	m.year	date	date
Resource Type			object type	doc type	doc type	type	type
Format					mime type	format	format
Resource ID							
Source							
Language	society name language name			language	language	language	language content.language
Relation							
Coverage Time		date	date period	date year	m.date m.year	coverage.t	date
Coverage Location	Continent Country Ethnic Region	cultural region geo region	location content place	location		coverage.l	Continent Country Region
Rights							

For almost all metadata sets it makes sense to describe the location with which the resource is primarily associated.

- In NECEP the area is described where the society is located, i.e. also related objects such as images, videos etc are associated with that geographical area. The information is contained in three levels of detail.
- In the RMV catalogue the aerial information is contained in two fields “cultural region” and “geographic region”. The cultural region is ambiguous since in many cases ethnic information will be mentioned.
- The Fotothek has two entries that could map. They have an element “location” that contains information about the place of creation. The element “content place” refers to a place that is referred to in the document itself (a painting created in Rome can include a scene from Egypt).
- The IMDI set used in the languages domain elements that refer to the geographical area in three levels.
- DC has the field coverage that has a qualifier for aerial coverage.

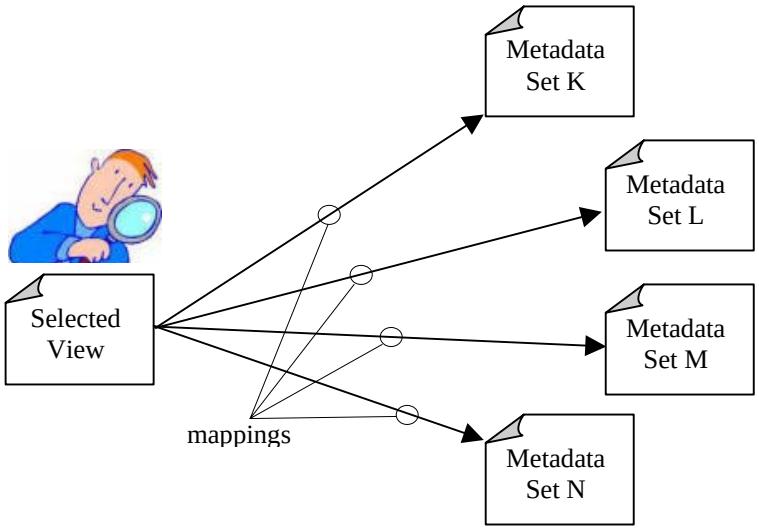
The elements that contain language information have two different meanings, they can refer to the language a document is about or a language a document is in. So a text can be in English, but describe the Trumai language. Different user groups are interested in different aspects of this.

- DC’s language field has the meaning “the language a document is written in”. One would describe the language a document is about in the “subject” element. Yet there is no qualifier for this, so we don’t know whether the element is used to encode this.
- NECEP has a language element, but it also has a society element. Often the language and society names are the same or at least similar.
- The HoS-Berlin set has the element “language” but it is assumed that they only code the language a document is written in.
- The IMDI set is specialized and has options for both.

In fact we can’t differentiate between the two meanings at the beginning.

The most difficult element (element sub-set) is the content description. Completely different dimensions and thesauri are used for content encoding.

- DC uses the element subject which is however not specified in more detail. So it can include all types of content description values.
- The NECEP set is meant to describe societies, so the society is the object. In this way almost all elements describe the content.
- The RMV catalogue has an element called categorization. The value this element can take is a list of keywords extracted from the AAT thesaurus (see appendix A). So basically the content description has one dimension filled with keywords classifying a given object.
- The Fotothek uses primarily two entries "primary iconography" and "secondary iconography". Both elements can have values that are taken from the complex IconClass thesaurus (see appendix D). The construction is similar to that one of RMV, however, the classes differ considerably.
- The HoS Berlin archive has in its metadata sets the element "keywords", but they are not yet specified.
- The IMDI set has a rather elaborated sub-set to describe the content. The sub-elements are Genre, SubGenre, CommunicationContext, Task, Modality, Subject, Description and Keys[7]. Task and Subject both of which are fairly unconstrained can be mapped most easily with what other domains describe as content.



Special concern has to be devoted to the question of how to map the content descriptions to allow useful joint queries. We first have to check how these elements are actually used within the domains. A careful analysis may reduce the necessary effort.

Summarizing we can say that only a start with pair wise comparison lead us to useful interpretations (see appendix G). From these we will derive per view mappings to all other sets as indicated in the above figure. We realize also that at this moment we start from the proper definitions of the semantics of the elements. However, it is known that the usage of the elements varies to a certain extent, i.e. for the second phase we will have to check the usage of elements.

2.3 Formal Framework for Mapping

The mapping requires a number of information types:

- definition of terms in English (element names, controlled vocabulary elements)

[7] The Language element, describing the language the resource is about, is also part of the content description block.

- dedications of all terms to the following languages:
    - French
    - German
    - Italian
    - Swedish
    - Dutch
- the relations between the terms
- alternatives (synonyms) in some cases as for language and society names

Alternatives are seen as special type of relations.

All definitions will appear in the DORA namespace for matters of simplicity, although the IMDI definitions are currently being integrated in open RDF-based repositories.

For the term definitions we will use the following schema[8]:

termID
term-name
term-XPath
domain-name
sub-domain-name
dedications
    fra = french-name
    ger = german-name
    ita = italian-name
    swe = swedish-name
    dut = dutch-name

For the relations we will use the following schema:

namespace:termID
    namespace:termID
    relation-type
    match-factor

The terms can be elements of the metadata sets, but also elements of the controlled vocabularies of elements. In some cases thesauri are used. It has to be analyzed yet in how far an equality of nodes in such thesauri implies an equality of sub-trees.

Within the project we have to find out what kind of relation types will be used. At first instance we will make use of the “equality” relationship from OWL and define a “maps_to” relationship. This relationship is associated with a matching factor that specifies the degree of match between 1 and 3 with “1” meaning an almost perfect match. This can be used while searching as an indicator of how much noise is expected. It could also be used for ranking.

A deeper semantic modeling could be carried out, but this would require more time and specialists. Therefore, we will not include this in the current ECHO project. Therefore, also we are not interested in specifying everything in RDF right now. We will use a specific search engine that makes use of the simple relation types. The schemas for the two structures can be found in appendix K.

3. Access to Resources

(this will become available in the second version of this document)

[8] The schemas will be translated to XML/RDF schemas within the first phase implementation.

Appendix A

Metadata set used by the RMV

The following elements are used within the Ethnology Museum in Leiden (RMV).

Nr	Element Name	Description	mapping
1	Cultural Origin (DCCoverage temporal)	• Culture, style and period taken from the OMV thesaurus, which is continent and region oriented • Religion oriented description (society, ...)	st
2	Date (DCDate)	different formal options are given: exact date dd-mm-yyyy from/to yyyy/yyyy before yyyy after yyyy about yyyy before 00 yyyy (vC)/yyyy (vC)	st
3	Presentation Title (DCTitle)	short title to be used in exhibitions; there can be other title choices such as: sorting title, local title, official title, series title, descriptive title, printing title, function title, English title; there is a field to specify the language the title is in	pr
4	Name of Object (DCTitle)	short but specific object indication ; additional information can be associated such as sorting name, alternative name, active name; also here the language can be specified	pr
5	Material/Fabrication	a description of the major materials the object exists of; can be several terms	-
6	Size	physical size of object	-
7	Special Physical Features	possibility to indicate special features of the object	-
8	Publicly oriented Description (DCCustom1)	a prose description of the object that can be used for public presentations	pr
9	Object History	this element offers the possibility to mention the collection the object was part of beforehand or a number that identifies its relation to an earlier exhibition or so	-
10	Geographic Origin (DCCoverage spatial)	location where the object was used; all geographic terms have to be taken from the OMV thesaurus; some additional info can be specified such as sorting location, comments	st
11	Categorization (DCSubject and keywords)	description of the content with the help of keywords extracted from the OMV category thesaurus;	st
12	Source Links	references to different types of sources such as publications, related literature, unpublished documents, exhibitions; for each of these there is a field	-
13	Reference to Digital Object	not yet fully defined	-
14	Others	not yet fully defined, manual speaks about meta objects	-

For mapping purposes we can identify three different options: no usage (-), usage in a structured way (st), usage as free prose text (pr).

Content Description

The content is described by categories according to the SNVT thesaurus. Here we want to introduce the main categories and discuss their usefulness for the joint infrastructure.

Nr	Category	mapping to HoA	mapping to HoS	mapping to languages
01	Chasing, fishing & gathering	can have similar motives encoded in IconClass and texts	can have similar motivs encoded in texts or titles	can have similar motives encoded in texts or in MD content
0101	chasing			
0102	fishing			
0103	gathering			
02	weapons & war	can have similar motives encoded in IconClass and texts	can have similar motivs encoded in texts or titles	can have similar motives encoded in texts or in MD content
0201	hand weapons & accessories			
0202	attack weapons & accessories			
0203	defense and protection means			
0204	gallery weapons			
0205	artifacts related with war			
03	farming, gardening & foresting	overlap little	overlap little	can have similar motives encoded in texts or in MD content
0301	farming & gardening			
0302	foresting			
04	veeteelt	overlap little	overlap little	overlap little
0401	vee en pluimvee houding			
0402	insekten teelt			
05	food, drank& consumption	overlap little	overlap little	overlap little
0501	preparation of voedsel			
0503	type of voedsel			
0504	drank			
0505	opdienen en consumeren			
0506	conservation and storage			
0507	genotmiddelen			
06	clothing & verziering	overlap little	overlap little	overlap little
0601	cloths			
0602	shoes			
0603	bewerking van her lichaam			
0604	sieraden			
0605	accessoires			
07	Lichaamsverzorging, geneeskunde, perzoonlijk comfort	overlap little	overlap little	can have similar motives encoded in texts or in MD content
0701	lichaamsverzorging			
0702	Medical care			
0703	perzoonlijk comfort			
08	housing	can have similar motives encoded in IconClass and texts	can have similar motivs encoded in texts or titles	can have similar motives encoded in texts or in MD content
0801	preparation of housing			
0802	parts of construction			
0803	huisraad			
0804	light, heating, vuur			
0805	house animals			
0806	water supply			
0807	constructions			
09	nijverheid, trade & services	overlap little	can have similar motivs encoded in texts or titles	
0901	gathering of grondstoffen & natuur produkten			
0902	materialbewerking			
0903	industry			
0904	second hand usage			
0905	size and weight			
0906	ruilmiddelen			
0907	trade and services			

10	vervoer	overlap little	can have similar motivs encoded in texts or titles	overlap little
1001	vervoer door mens/mechanica			
1002	vervoer door dieren/mechanic			
1003	vervoer over water			
1004	routes en hulpmiddelen			
1005	vervoer door de lucht			
11	communication	overlap little	overlap little	overlap little
1101	mnemotechnical means			
1102	scripts			
1103	signaling means			
1104	education, onderricht, leermiddelen			
1105	aanschouwelijk maken, explicatie, overdracht			
12	social, political, juridical	overlap little	overlap little	overlap little
1201	status, rang, waardigheidstekenen			
1202	law system			
1203	artifacts related to slavery			
1204	memorabilia			
13	life cycle	overlap little	overlap little	can have similar motives encoded in texts or in MD content
1301	zwangerschap, geboorte en eerste levensjaar			
1302	initiatie			
1303	huwelijk			
1304	ouder worden			
1305	dood en rouw			
14	religion and ritual	overlap little	overlap little	can have similar motives encoded in texts or in MD content
1401	representations van het bovenaardse			
1402	cultus en andere heilige objecten			
1403	altaren, heilige ruimten en aankleding			
1404	offers, offeranden en offergereedschap			
1405	magische bescherming en afweer			
1406	hulpmiddelen bij ritueel, gebed, meditatie, waarzeggerij en magie			
1407	symbols of religious state			
15	arts	overlap little	overlap little	can have similar motives encoded in texts or in MD content
1501	dans, dansornaat en toebehoren			
1502	toneelspel			
1503	beeldende kunst			
1504	cartografie			
1505	music			
16	relaxing, sports, games	overlap little	overlap little	overlap little
1601	toys for children			
1602	equipment for sports and games			
1603	hebbedingetjes, verzamelobjekten			
17	others	overlap little	overlap little	overlap little
1701	others			
1702	onbepaalt vaatwerk			
1703	onbepaalt textiel			

The object is classified according to these categories, i.e. a set of numbers determines what this object is. For some categories there are even more fine-grained semantics that seem to be difficult to use in an interoperable scenario.

Meaning of classification: If an object falls into the categories 0205 and 1505 then we may conclude that the object is a song about war. When further the cultural origin says that the object

is from the Amazonas area in Brazil we may find it if someone searches for music related to war for the Trumai people (a tribe living in the Amazonas area).

Appendix B

Metadata set used by in the History of Science (Berlin)

The metadata set such as recently proposed by the HoS group is primarily focusing on management tasks, i.e. the amount of elements that describe the content of a resource is small. The set is a flat list that offers a category “meta” that can be used to enter Dublin Core type of descriptions.

element	sub-element	comment
description		informal textual description of the resource
name		filename of the resource
creator		project or person that created the resource
archive-creation-date		time and date of creation of the archive entry
archive-storage-date		not useful within DORA
archive-path		
derive-from		
	archive-path	
	description	
linked-with		
	archive-path	
	description	
content-type		document type comparable to MIME type
meta		substructure see below
dir		not useful within DORA
	description	
	name	
	path	
	meta	substructure see below
file		not useful within DORA
	description	
	name	
	path	
	date	
	modification-date	
	creation-date	
	size	
	mime-type	
	md5cs	
	meta	substructure see below

The meta substructure contains elements that are partly dependent on the type of document. The generic type as listed in the following may give an impression.

language		the language a document is in
DRI		not useful for searching
context		
	link	link to collection as a context
	name	description of that collection
generic		
	author	Dublin-Core type of fields
	year	
	title	
	secondary-author	
	secondary-title	
	volume	not useful for searching

	number	
	pages	
	date	Dublin-Core type of field
	place-published	not useful for searching
	publisher	
	edition	
	tertiary-author	DC type of fields
	tertiary-title	
	number-of-volumes	not useful for searching
	type-of-work	
	subsidiary author	
	alternative-title	
	isbn-issn	
	call-number	
	label	
	keywords	useful but unconstrained
	abstract	
	notes	not useful for searching
	url	

Appendix C

Metadata set used by in the History of Science (IMSS)

Here we will list the elements used for describing instruments. The other two schemes for documents and artistic objects share the same core and are very similar.

element	comment
belongsTo	not useful for searching
contextualized	not useful for searching
DCcontributor	name of artists or engineers
DCcopyright	not useful for searching
DCcoverage	not yet clear how the field will be used
DCcreator	name of artists etc
DCdate	
DCdescription	prose text
DCformat	not yet clear how the field will be used
DCidentifier	not useful for searching
DClanguage	to describe the language the descriptions are in
DCpublisher	not useful for searching
DCrelation	not useful for searching
DCsource	not useful for searching
DCsubject	not yet clear how the field will be used
DCtitle	
DCtype	not yet clear how the field will be used
Giver	not useful for searching
hasComponentType	not useful for searching
hasInstrumentType	not useful for searching
hasWR	not useful for searching
historicallyLocatedIn	not useful for searching
inventor	?
isDedicated	not useful for searching
isDocumentedIn	not useful for searching
isPartOf	not useful for searching
locatedIn	not useful for searching
objectRelated	not useful for searching
owner	not useful for searching
preservedIn	not useful for searching
purchaser	not useful for searching
receiver	not useful for searching
refersToDiscipline	not useful for searching
relatedConcept	not useful for searching
shortname	not clear whether useful
shown	not useful for searching
simulatedBy	not useful for searching
usedFor	not useful for searching
user	not useful for searching

IMSS uses a flat list where a number of pointers contain relations, i.e. implicitly a hierarchical scheme is realized. For us it is not clear yet for all fields how they will be used. Examples will help.

Appendix D

Metadata set used in the Fotothek (Biblioteka Herziana Rom)

For the Fotothek BH uses the MIDAS rules to describe their image objects with metadata records. The purpose of the MIDAS rules is beyond the pure discovery and is also used for management. It is a fairly exhaustive structured description set and allows creating linked hierarchies between objects. Only the most relevant elements are shown in the following table. The important description of the content of an image is done according to the IconClass rules.

Object-Document		Obj	Description
Objekt-Verwalter		ob28	description fields about owner or administrator
	Ort	2864	
	Verwalterart	2890	
	Name-Museum	2900	
	Abteilung	2930	
	Inventar-Nr	2950	
	Person	2910	
	Titel	2914	
Objekt-Aufbewahrung			description fields about where the object is housed: some geographical or topographical information like Australia, Venice
	Ort	5108	
	Ortsteil	5110	
	Straße	5116	
	Nr	5117	
	Stelle	5125	
Objekt-Künstler		ob30	description fields about artist
	Name	3100	
	Name in BH	31bh	
	Authentizität	3470	
	Tätigkeit	3475	
Datierung		5064	date of creation
Zeitangabe		5062	could be any other date descr.
Entstehungsort		5130	place of creation here “Kunststil” like Venetian etc…
Objekttitel		5200	known name of the object
Bauwerkname		5202	instead of 5200 for building
Gattung		5220	sub-genre for paintings
Art		5226	topic of sub-genre
Sachbegriff		5230	Object type
Material		5260	type of material used
Technik		5300	type of technique used
prim. Ikonogr.		5500	primary content descr
sec. Ikonogr		5510	secondary content descr
lokaler Bezug		5560	place the content refers to
Objekt-Bauwerk		ob26	Description of the relation between the object and a building (there are many more descriptive fields)
	Ort	2664	
	Sachbegriff	2690	
	Träger	2694	
	etc		
Objekt-Person		ob40	Relation to other person
	Name	4100	
Beziehung zu Objekt		5007	Relation to other object and description of other object (a normalization would be better, i.e. to include the object as a regular one in the domain and have just a link to it)
	Link	5008	
	Hersteller	5009	
	Sachbegriff	5010	
	Titel	5013	

	Bauwerk	5014	
	Ort	5015	
	Zeit	5011	
	etc		
Kurztitel		7910	
Literaturnachweis		8350	
Foto		8450	Description of the photo of the
	Nummer	8470	
	Verwalter	8460	
	Fotograf	8490	
	AufnahmeDatum	8498	
	Zugangsdatum	8496	
	Inhalt	8510	
	Signatur	8515	
	Dateiname	8540	
	Kommentar	9990	
	Urheber	9902	
	etc		

The content is described according to the IconClass proposal that is widely used in the arts domain. IconClass was worked out by Dutch scientists and is available at the Dutch academy of sciences.

(a short description will follow – the thesaurus is too large to be described fully at this place)

Appendix E

Metadata set used in the Lineamenta Project (Biblioteka Herziana)

The Lineamenta collection uses internally a rich description set, however, it seems that they will only export a limited set. For this export the same core metadata set is used as for the History of Science – Berlin collections. They use a slightly different specialized “meta” set that is indicated here.

element	comment	DORA usage
image	reference to an image	not useful for DORA search
language	language the document is written in	useful
document type	associated with fixed vocabulary	useful
title		useful
person	equivalent to DC:creator and contributor	useful
location	detailed spec where the object is placed	useful
date		useful
object	detailed description of the object	useful
keywords	not clear how used	?

Here further examples should be made available.

Appendix F

Metadata set used in the Maps of Rome Project (Biblioteka Herziana)

The descriptive data is kept in a relational database that has three tables: PDR, Piantecopie, Persone. These were exported to separate XML documents.

From these XML documents received we can identify the following metadata elements that are relevant for DORA:

element	comment	DORA usage
author-name	metadata elements describing the author	useful
alternative names		useful
date of birth		not useful
date of deadth		not useful
place of birth		not useful
place of acting		not useful
date		useful
title		useful
method	not clear whether this can be mapped	?
dim-alt		not useful for searching
dim-long		not useful for searching
orientation		not useful for searching
incisore	is it a relevant contributor?	?
editor		useful
huelsen	these terms are not yet clear	?
scaccia		?
frutaz		?
rome-veduta		?
description		useful
collection	probably not a search term at DORA level	?
image reference		for backlinking

This list has to be checked with Bibl Hertziana.

Appendix G

Metadata set used in the Language Domain

All metadata descriptions in the language area are created according to the IMDI standard (see www.mpi.nl/IMDI). IMDI provides a structured set that is used for resource discovery and management.

Session		
Name		
Title		
Date		
Location		
	Continent	
	Country	
	Region +	
	Address	
Description +		
Resource Reference		
Keys		
Project +		
	Name	
	Title	
	Id	
	Contact	
	Decription +	
Content		
	Genre	
	SubGenre +	
	Communication Context	
		Interactivity
		Planning Type
		Involvement
		Social Context
		Event Structure
		Channel
	Task	
	Modalities	
	Subject +	
	Languages	
		Language +
		Description +
	Description +	
	Keys	
Actors		
	Description +	
	Actor +	
		Resource Refs
		Role
		Family Social Role
		Name +
		Full Name
		Code
		Language +
		Ethnic group
		Age
		Sex
		Education
		Anonymous
		Contact
		Description +
		Keys

Session		
Resources		
	Media File +	
		Resource Id
		Resource Link
		Type
		Size
		Format
		Quality
		Recording Conditions
		Position
		Access
		Description +
	Written Resource +	
		Resource Id
		Resource Link
		Media Resource Link
		Date
		Type
		SubType
		Format
		Size
		Derivation
		Content Encoding
		Character Encoding
		Validation
		Access
		Language Id
		Anonymized
		Description +
	Source +	
		Id
		Format
		Quality
		Position
		Access
		Description +
	Anonyms	
		Resource Link
		Access
References		
	Description +	

Language	
	Id (ccv)
	Name + (str)
	MotherTongue (ccv)
	Primary (ccv)
	Dominant (ccv)
	Description + (sub)

Keys	
	Key + (sub)

Key	
	Name = Value (string)
	Vocabulary Link (c)

Description	
	Text (string)
	Language Id (ccv)
	Link (c)
	Name (string)

Validation	
	Type
	Methodology
	Level
	Description (sub)

Access	
	Availability (string)
	Description + (sub)
	Date (c)
	Owner (string)
	Publisher (string)
	Contact (sub)

Contact	
	Name (string)
	Address (string)
	E-mail (c)
	Organisation (string)

Resource Reference	
	Type (cv)
	SubType (ocv)
	Format (cv)
	Link (c)

Appendix H

Metadata set used Philosophy

This section will be added soon.

Appendix I

Dual Mapping between Structured Elements

This chapter can be seen as exercises to come to final mappings for the different views (see appendix J). For a couple of dual sets some topics are discussed that are relevant and indicate the problems that we expect.

The NECEP-RMV mapping makes sense since NECEP describes societies in detail of which RMV will have objects in its repository.

NECEP	RMV	comment
A1 society names	cultural region	has to be checked whether values are the same, probably value matching necessary
A7 alternative names	cultural region	has to be checked whether values are the same, probably value matching necessary
B? continent	cultural region geographical region	RMV has two fields that apply, details have to be checked
B1 country	cultural region geographical region	RMV has two fields that apply, details have to be checked
B3 ethnic region	cultural region geographical region	RMV has two fields that apply, details have to be checked
C1 language name	cultural region	a mapping between languages and societies is necessary

The NECEP-IMDI mapping makes sense since NECEP describes societies for which one can probably find language resources in the languages domain.

NECEP	IMDI	comment
A1 society Names	language name	a mapping between languages and societies is necessary
A7 alternative names	language name	
B? continent	continent	perhaps mapping due to different names
B1 country	country	perhaps mapping due to different names
B3 ethnic Region	region	perhaps mapping due to different names
C1 language name	language name	perhaps mapping due to different names

The RMV-IMDI mapping makes sense since one may find objects in the RMV repository that may be related with language resources.

RMV	IMDI	comment
fields mentioned above will be used	see above	
date	date	rmv.date is date of creation; imdi.date is date of recording; overlap seems to be small
categorization	content	rmv.categorization contains a set of numbers describing the type of content included; IMDI uses a whole sub-structure for content; has to be checked how this can be mapped

With respect to the HOS-IMDI mapping we don't expect too much overlap in the scope of the ECHO project. There may be language resources that appear in both repositories.

HoS Berlin	IMDI	comment
creator	actor	not much overlap to be expected
meta.author[9]	actor	not much overlap to be expected
language	language	here is a difference: hos.language refers to the language the resource is in while imdi.language refers to the language the resource is about; nevertheless, hos.language could be useful for linguists;
meta.year	date	hos.meta.date means year of publication while imdi.date refers to the date of the recording
title[10]	content	

[9] The hos set includes secondary and tertiary authors. The indicated mapping should include them as well.

	title	
keywords	content	hos.meta.keywords describe the content of the resource and can be mapped with the content description in IMDI; it is not clear how keywords will be used in HoS

With respect to the IMSS – IMDI mapping we don't expect too much overlap as well despite the formal overlap between the fields used.

HoS IMSS	IMDI	comment
DCcontributor	actor	
DCcoverage	location, date	IMSS will have to use qualifiers to separate the two information types
DCcreator	actor	
DCdate	date	
DCformat		
DClanguage	language	in IMSS probably the language the document is in, in IMDI both is possible
DCsubject	content	no information yet how this field will be used
DCtitle	title	
DCtype	type	
inventor	actor	not yet clear whether this field is relevant

In the current ECHO project we do not expect too much overlap, which is due to the fact that both repositories will not have too many resources that are related. However, in principle much overlap can be expected, since texts from the language resource area can for example explain objects in the HoA area.

HoArts Fotothek	IMDI	comment
3100 name artist	actor	overlap estimated to be small
5064 date	date	hoa.date is precise; hoa.period offers different options; both can be matched with imdi.date
5062 period	date	
5130 location of creation	location	
5200 object title	title	
5202 title of building	title	hoa title in case of buildings
5230 object type	content	not yet clear whether there is a potential for matching
5500 prim iconography	content	here a classification according to the IconClass system is used
5510 sec iconography	content	
5560 place of content	location	location as part of the content of the painting

Not much overlap is expected since the resources probably are not that much related.

HoArts Lineamenta	IMDI	comment
document type		no real equivalence in IMDI since the vocabulary is different
creator	actor	overlap estimated to be small
m.language	language	Lin is encoding the language the document is in
m.person	actor	overlap estimated to be small
m.year	date	
m.title	title	
m.date	date	
m.keywords	content	no specifications yet as how to fill in keywords
object	title	
m.location	location	in Lin no formal distinction in continent, countries etc

Here one can expect some overlap in principle. However, the metadata set chosen by HoS does not allow to draw too many relations.

[10] The HoS set includes secondary and tertiary titles. The indicated mapping should include them as well.

HoArts Fotothek	HoS Berlin	comment
3100 name artist	creator meta.author	
5064 date	meta.year	
5062 period	meta.year	
5200 object title	title(s)	
5202 title of building	title(s)	
5230 object type	keywords	it is not yet clear how keywords will be used in HoS
5500 prim iconography	keywords	it is not yet clear how keywords will be used in HoS
5510 sec iconography	keywords	it is not yet clear how keywords will be used in HoS

A number of Dublin Core mappings will be used. Therefore, we compare some sets from the DC view point.

Dublin Core	HoS-Berlin	comment
DCcontributor	author secondary author tertiary author	
DCcoverage	year	
DCcreator	author secondary author tertiary author	
DCdate	date	
DCformat	document type mime type	
DClanguage	language	
DCsubject	keywords	
DCtitle	title secondary title tertiary title	
DCtype	doc type	DC not very clear – so not clear how to map

The mapping between DC and IMDI is fairly straightforward.

Dublin Core	IMDI	comment
DCcontributor	participant	
DCcoverage	location date	
DCcreator	participant	
DCdate	date	
DCformat	format	
DClanguage	language	DC language is language a document is written in
DCsubject	content language	not at all clear how subject is used language the doc is about would fall under DC:subject
DCtitle	title	
DCtype		DC semantics not very clear

The mapping between DC and HoA-Fotothek.

Dublin Core	HoA-Fotothek	comment
DCcontributor	3100 name artist	
DCcoverage	5062 period 5130 place	
DCcreator	3100 name artist	
DCdate	5064 date	
DCformat		
DClanguage		
DCsubject	prim iconography sec iconography 5220	not at all clear how subject is used

DCtitle	5200 object title 5202 building title	
DCtype	object type	DC semantics not very clear

The mapping between RMV and DC does not give many options.

Dublin Core	RMV	comment
DCcontributor		
DCcoverage	date geographic origin	
DCcreator		
DCdate	date	
DCformat		
DClanguage		
DCsubject	categorization	
DCtitle	presentation title name of object	
DCtype		

Appendix J

Mapping for Views

As mentioned above we have to evaluate the usage of the various fields to optimize the mapping schemes. First it seems to be handy to describe the metadata elements to be used in short form as an overview.

Set	element name	appearance
NECEP	society names	society names
	alt soc names	society names
	continent	continent
	country	country
	ethnic region	ethnic region
	language name	language name

Set	element name	appearance
RMV Leiden	cultural origin	cultural region
	geographic origin	geographic region
	date	date
	categorization	content categories
	presentation title	presentation title
	object name	object name

Set	element name	appearance
Fotothek	artist (3100)	artist
	date (5064)	date
	period (5062)	period
	location (5130)	location
	object title (5200/02)	object title
	object type (5230)	object type
	iconography prim	iconography
	iconography sec	iconography

Set	element name	appearance
Lineamenta	continent	continent
	country	country
	title	title
	creator	creator
	date	date
	object	object
	content type	content type
	document type	document type
	language	language
	keywords	keywords

this set is derived from the XML files we received

Set	element name	appearance
IMSS	object type	object type
	creator	creator
	date	date
	subject	subject
	title	title
	type	type
	format	format
	language	language
	contributor	creator
	inventor	creator
	coverage.loc	location
	coverage.time	time

Set	element name	appearance
IMDI	language	language
	continent	continent
	country	country
	region	region
	date	date
	subject	subject
	task	task
	genre	genre
	subgenre	subgenre
	actor role	actor role
	actor name	actor name
	title	title

the marked fields describe the content

Set	element name	appearance
HoS Berlin	creator	author
	author	author
	sec author	author
	tert author	author
	language	language
	year	year
	title	title
	sec title	title
	tert title	title
	keywords	content keywords

Set	element name	appearance
Rome Maps	author-name	author
	altern author	author
	date	date
	title	title
	editor	editor
	location=”rome”	location=”rome”

Set	element name	appearance
Philosophy		

1. DC View

We refer to the names in the table above.

DC	Ethnology		History of Arts			History of Science		Philosophy	Languages
	NECEP	RMV	Fotothek	Lineamenta	Rome Maps	Berlin	IMSS		IMDI
Title		object name present. title	object title title of building	title	title	title	title		title
Creator			artist name	person	author-name editor	creator author	creator		actors
Contributor			artist name	person	author-name editor	creator m.author	contributor		actors
Subject		categorization	prim icono sec icono	object keywords	“rome maps”	keywords	subject		content
Date		date	date period	date	date	year date	date		date
Resource Type			object type	document-type		content-type	type		type
Format	“jpg”, “mpeg”, “wav”	“jpg”	“jpg”	“tiff”, “jpg”	“jpg” “image”	mime type	format		format
Language	society name language name			language		language	language		language content.language
Coverage Time		date	date period	date year	date	date year	coverage time		date
Coverage Location	Continent Country Ethnic Region	cultural region geo region	institute place place of creation content place	location	rome-veduta “rome”		coverage loc.		Continent Country Region

2. Necep View

NECEP	Ethnology		History of Arts			History of Science		Philosophy	Languages
	NECEP	RMV	Fotothek	Lineamenta	Rome Maps	Berlin	IMSS		IMDI
society names		cultural origin							language
alt soc names		cultural origin							language
continent		cultural origin geo origin	institute place place of creation content place	location	“europe”				continent
country		cultural origin geo origin	institute place place of creation content place	location	“italy”				country
ethnic region		cultural origin geo origin	institute place place of creation content place	location	“rome”				region
language name		cultural origin				language	coverage.l		language

3. RMV View

RMV	Ethnology		History of Arts			History of Science		Philosophy	Languages
	NECEP	RMV	Fotothek	Lineamenta	Rome Maps	Berlin	IMSS		IMDI
cultural origin	society names continent country regions		institute place place of creation content place	location	“europe” “italy” “rome”		coverage.loc		language continent country region
date			date period	date	date	year	coverage.time		date
present.title			object title	title object	title	title	title		title
object name			object title	title object	“rome maps”		title		title
geo origin	continent country regions		institute place place of creation content place	location	“europe” “italy” “rome”		coverage.loc		continent country region
categorization			prim.iconogr.	keywords		keywords	subject		content

			sec. iconogr.						
custom1				description	description	description			description

4. Fotothek View

Fotothek	Ethnology		History of Arts			History of Science		Philosophy	Languages
	NECEP	RMV	Fotothek	Lineamenta	Rome Maps	Berlin	IMSS		IMDI
institute	continent country region	cultural origin geo origin		location	“europe” “italy” “rome”		coverage.loc		continent country region
place	continent country region	cultural origin geo origin		location	“europe” “italy” “rome”		coverage.loc		continent country region
place of creation	continent country region	cultural origin geo origin		location	“europe” “italy” “rome”		coverage.loc		continent country region
content place	continent country region	cultural origin geo origin		location	“europe” “italy” “rome”		coverage.loc		continent country region
object title		object name presentat. title		title object	title	title	title		title
building title				object					
short-title				title object					
artist name				title creator person	author-name editor incislink	creator author	creator		actors
artist				creator					
person name				person	editor incislink author-name				
date		date		date	date	year date	date coverage.t		date
period		date		date	date	year date	date coverage.t		date
type				document type					
object type				document type			type		
prim. iconogr.		categorization		keywords	“maps”	keywords	subject		content
sec. iconogr.		categorization		keywords	“maps”	keywords	subject		content

5. Lineamenta View

Lineamenta	Ethnology		History of Arts			History of Science		Philosophy	Languages
	NECEP	RMV	Fotothek	Lineamenta	Rome Maps	Berlin	IMSS		IMDI
location	continent country region	geo. origin cultural origin	place of creation content place place institute		“europe” “italy”		coverage.loc		continent country region
title		object name present. title	object title artist name short-title		title	title	title		title
creator			artist name artist		author-name alternat. names	creator author	creator contributor inventor		actors
date		date	date period		date	date year	date coverage.time		date
object			object title building title short-title		“rome maps”				title
content type			object type		“maps”				
document type			type		“printed map” “landscape drawing”		type		
language	language name	cultural origin			“italien”	language			content.language
keywords		categorization	prim.iconogr. sec. iconogr. object type		“maps”	keywords	subject		language
description		custom1			description	description			description
person			artist name person name		editor engraver author-name				

6. HoS Berlin View

HoS Berlin	Ethnology		History of Arts			History of Science		Philosophy	Languages
	NECEP	RMV	Fotothek	Lineamenta	Rome Maps	Berlin	IMSS		IMDI
author			artist	creator	author-name editor		creator		actors
language	language name society names	cultural origin					language		language
year		date	date period	date	date		date coverage.time		date
date		date	date period	date	date		date coverage.time		date
title		present. title object name	object title	title object	title		title		title
description		custom1		description	description				description
keywords		categorization	prim.iconogr sec.iconogr.	keywords	“maps”		subject		content

7. IMSS View

(same as the DC view)

IMSS	Ethnology		History of Arts			History of Science		Philosophy	Languages
	NECEP	RMV	Fotothek	Lineamenta	Rome Maps	Berlin	IMSS		IMDI
Title		object name	object title title of building	title	title	title			title
Creator			artist	person	author-name editor	creator author			actors
Contributor			artist	person	author-name editor	creator author			actors
Subject		categorization	prim icono sec icono	object keywords	“rome maps”	keywords			content language
Date		date	date period	date	date	date year			date
Coverage Time		date	date period	date	date	date year			date
Type			object type	content-type		content-type			type
Format	“jpg”, “mpeg”, “wav”	“jpg”	“jpg”	“tiff”, “jpg”	“jpg” “image”	mime type			format
Language	society name language name			language		language			language content.language
Coverage Location	Continent Country Ethnic Region	cultural origin geo origin	place of creation content place place institute	location	rome-veduta “rome”				Continent Country Region

8. Language View

Language	Ethnology		History of Arts			History of Science		Philosophy	Languages
	NECEP	RMV	Fotothek	Lineamenta	Rome Maps	Berlin	IMSS		IMDI
language	society name language name	cultural origin		language		language	language		
continent	continent	cultural origin geo origin	place of creation content place place institute	location	“europe”		coverage.loc		
country	country	cultural origin geo origin	place of creation content place place institute	location	“italy”		coverage.loc		
region	ethnic region	cultural origin geo origin	place of creation content place place institute		“rome”		coverage.loc		
date		date	date period	date	date	date year	date coverage.time		
content		categorization	prim.iconogr sec.iconogr	keywords	“maps	keywords	subject		
task		categorization	prim.iconogr sec.iconogr				subject		
genre		categorization		content-type	“maps’		subject		
actors			artist	creator	author-name editor	author	creator		
description		custom1		description	description	description			
title		presentation title object name	object title	title object	title	title	title		

Appendix K

Schemas

Schema for Term Definitions

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs='http://www.w3.org/2001/XMLSchema'>
<xs:element name=”term”>
   <xs:complexType>
      <xs:sequence>
         <xs:element name="termID" type="xs:ID"/>
         <xs:element name="term-name" type="xs:string"/>
         <xs:element name=”xpath” type=”xs:URI”/>
         <xs:element name="domain-name" type="xs:string"/>
         <xs:element name="sub-domain-name" type="xs:string"/>
         <xs:element name="description" type="xs:string"/>
         <xs:element name="dedications">
            <xs:complexType>
               <xs:sequence>
                  <xs:element name="fra" type="xs:string"/>
                  <xs:element name="ger " type="xs:string"/>
                  <xs:element name="ita" type="xs:string"/>
                  <xs:element name="swe " type="xs:string"/>
                  <xs:element name="dut " type="xs:string"/>
               </xs:sequence>
            </xs:complexType>
         </xs:element>
      </xs:sequence>
   </xs:complexType>
</xs:schema>
```

Schema for relations

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs='http://www.w3.org/2001/XMLSchema'>
<ECHO:schema xmlns:xs=’http://www.mpi.nl/echo/schemas/ECHO-def-schema’>
   <xs:element name=”mapping”>
      <xs:complexType>
         <xs:sequence>
            <xs:element name="termID" type="xs:ID”/>
            <xs:element name="termID" type="xs:ID"/>
            <xs:element name="relation-type" type="xs:string"/>
            <xs:element name=”match-factor” type=”xs:integer”/>
         </xs:sequence>
      </xs:complexType>
   </xs:element>
</xs:schema>
```