

Language Resource Archive Utilization

Metadata Layer

IMDI is a distributed domain of linked IMDI-XML files.
It is trivial to setup an IMDI portal.
allows to combine local and central resources

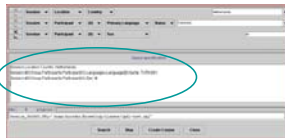
- browsing directly in the domain of linked IMID-XML files



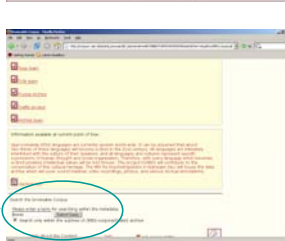
- browsing in HTML domain dynamically created via XSLT



- geographical browsing (soon to come)



- structured search with IMDI browser



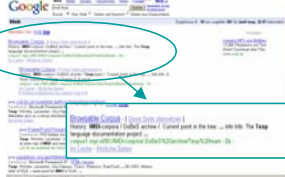
- structured search in HTML domain (soon to come)



- unstructured search with IMDI browser (soon to come)



- unstructured search in HTML domain



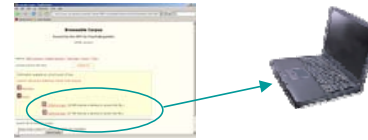
- DC/OLAC Gateway using OAI PMH

- Google Support (searching for IMDI Teop yields top node in archive)

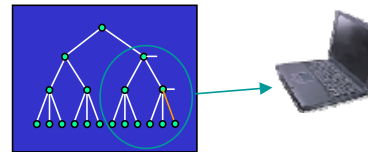
Content Layer

Important is that all resources are accessible in a few standard archival formats such as XML, CHAT, Shoebox, WAV, MPEGx. They should only be encapsulated internally by non-archival formats for fast indexes etc. Applications may exploit the bundling of resources such as for complex lexica that also include photos, sound and video clips. In doing so everyone can write his own tools for metadata and resource utilization – everything is well described.

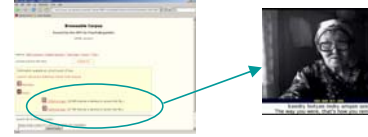
- IMDI browsing/searching, **downloading individual resources** and work on notebook



- IMDI browsing/searching, **downloading a whole subtree** inclusive resources and working on a local archive copy



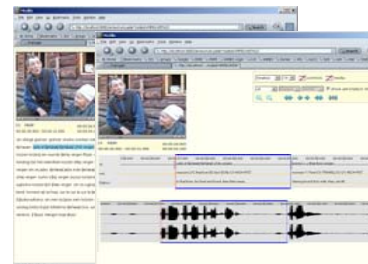
- IMDI browsing/searching and **direct playing** with some plug-in



- **web-based lexicon work:** select lexica from archive, start LEXUS, manipulate lexica and write back
 - LMF compliant (like LIEGO)
 - own work space
 - sharing a lexicon with others
 - merging of lexicon versions
 - IVE Shoebox, CHAT, XML
 - schema and entry change
 - connection with ISO DCR and MDF
 - connection to GOLD to come
 - search across different lexica
 - www.mpi.nl/lexus

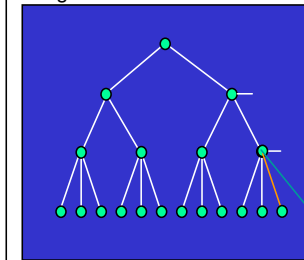


- **web-based annotation work:** select annotations from archive, start ANNEX and analyze (modification option to come)
 - ELAN like functionality via the web
 - fast search across many annotations
 - simple GUI intended
 - different viewing stereotypes
 - demo at <http://lux05.mpi.nl/ds/annex/>

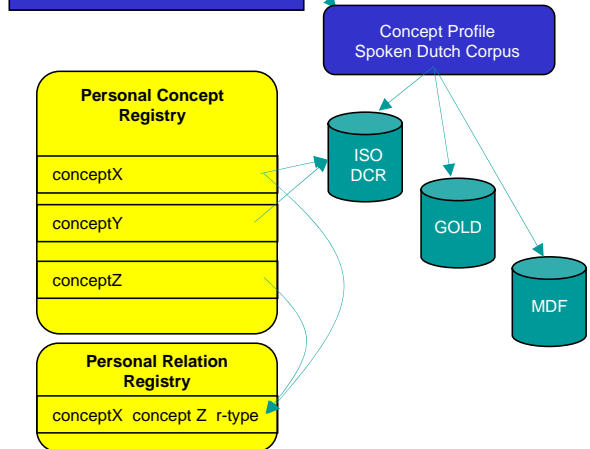


Ontology Support

LEXUS and ANNEX already support to search on different lexica and annotations, i.e. users must have a way to carry out mappings. The strategy MPI Nijmegen and MPI Leipzig are following includes both bottom-up and top-down concepts. Top-down concepts make use of "central" ontologies such as the ISO Data Category Registry and GOLD as reference frameworks. Bottom-up concepts allow users to easily define private concepts and mappings and register them.



Any single resource or a group of resources indicated by an archive node (typically representing a corpus) can be associated with a **concept profile** (CP). CPs can carry mappings to central ontologies such as for Spoken Dutch Corpus.



Users can easily create their own PCRs and PRRs, include concepts from CPs or define own ones that are used in individual resources, define personal mappings between them or point to central ontologies and create implicit mappings. Simple editors and browsers help to establish the personal registries and the search engine will make use of the mappings. In doing so we guarantee on the one side the need for individual researchers to flexibly react on their needs and on the other side we open the possibility of re-using definitions and mappings.

This will be implemented in 2005.