

Accessing Multimedia Language Resources the Easy Way

Daan Broeder, Peter Wittenburg

MPI
SDU
ISSCO
IMS



EAGLES/ISLE
Metadata Initiative
www.mpi.nl/ISLE



Documentation of
Endangered Languages
www.mpi.nl/DOBES



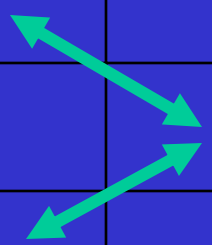
Max Planck Institute for Psycholinguistics

Resources & Tools

Relation of Mutual Dependency - Orthogonal

Dream: locate resource easily and directly execute a feasible tool on it independent of X, Y, Z

	Tool Registry	Resource Registry	Advice Registry
ACL/DFKI	+		
IMDI/MPI		+	
OLAC/LDC	+	+	+
others	?	?	?



Dream realistic?

Yes!!

Some teams are working on the problem!

see MIME Types + Standard Browsers

IMDI Goals:

- Search & Browse for (MM) Language Resources
by using a space defined by domain concepts (Metadata)
- When found one or a group of resources, indicate tools
which work on resources, and execute one of them
- Need much flexibility and choice for the user



What is IMDI?

- ISLE-based European Initiative
 - Metadata set primarily for MM LR
 - Tools Set (Editor, Browser, Search, Admin)
 - Showcase
- Deliverables (www.mpi.nl/ISLE)

IMDI

Demo

Demo

Done	To Do
White Paper (March 00)	Lexicon MD set
Metadata Overview	Mapping to DC/OLAC
Session + Catalogue MD set	Search Fct
Schemas for both	Editor (Lexicon), Scripts
Editor, Browser, Scripts	Infrastructure
Ref Card, Flyer	Interaction with SALT (terminology)
Demo (Showcase), sc tool connection	Tool Connection - large scale



ISLE

ISLE Metadata Initiative

Some Landmarks

- Library catalogue standards
- Dublin Core (originally librarians)
- Warwick Framework
 - concept of modular domain specific packages
- Resource Description Framework (RDF)
 - structure & semantics, reuse of packages
- *Discoverage of Chaos at MPI etc --> Browsable Corpus*
- DFKI Tool Registry
- March 00: IMDI White Paper
- Open Archives Initiative (OAI)
 - standardized metadata harvesting protocol, Ithaca TC
- December 00: Open Language Archives Community (OLAC)
- MPEG7 (large set also for media features etc)

DC



Brief Comparison

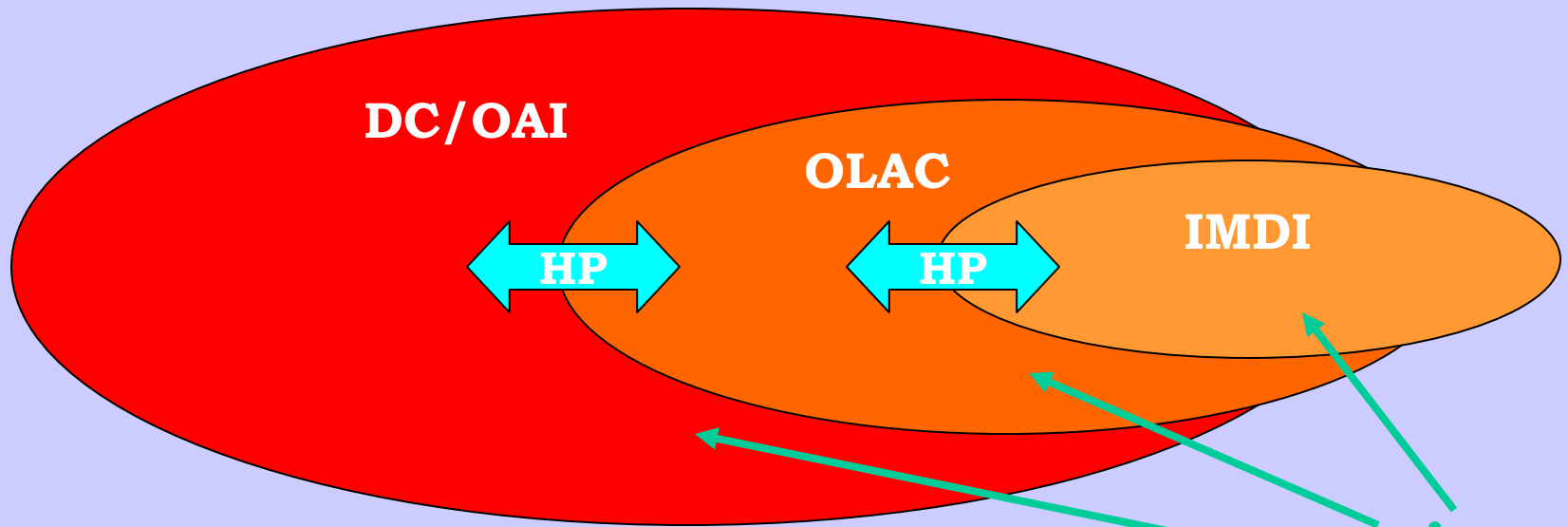
	DC	OLAC	IMDI
addressed community	world	linguists, S&L engineers	linguists, S&L engineers
scope	all sorts of Web-Resources	all sorts of Language Resources, Tools, Comments	mainly MM LR and Lexica
approach	top-down	top-down	bottom-up
operations	search	search	browse,search,exec
set	low-overhead	low-overhead	more detailed
topology	-	central harvesting	ch+local residence

Typical IMDI query:

Give me all MM resources where children of Australian aborigines were recorded both at the age of 6 and 8.



Metadata Mapping



Interoperability:

- is a must
- XML + mapping of elements and controlled vocabularies
MPEG7-DC harmonization very restrictive
- communication via OAI harvesting protocol
- “click through” for more detail or
- go beyond boundary with search



Few Practical Issues

- all is very new (especially to users) - will be faced with changes
- need a critical mass to see the gain - without gain no MD
- much work is manual work - need good tools
- efficiency is a big task - need handy scripts
reusage of blocks
not all elements necessary
iterative creation
- existing data - templates + scripts
- towards??



Resources & Tools I

How to get the two worlds together?

- OLAC now: store info of both - make switching easy
- IMDI: locate resource(s) and directly execute a tool

Required Information (all in MD set and at client):

- Resource Type (audio, video, text, photo, ...)
 - Format (CHAT, rDBMS, AIF, WAV, AVI, ...)
 - Encoding (UTF-8, MP3, MPEG, JPEG, ...)
 - Structure (for example for search)
 - Linguistic Encoding (trees, ...)
 - Operating Environment (CPU, OS - automatic extraction)
-) MIME Type
redundant?

MIME Types not sufficient



Resources & Tools II

Tools Registry

checks which tools fit

sends back info about tools

Resource Browser

determines info set for resource bundle
user specifies function

browser will download tool and start it
caching etc

will store user profiles
maintain history list

Demo

sends resource info

sends tool info

Need a Protocol

structured templates for information exchange in both directions

tool registry has to give service
tools need uniform call shell

bi-directional?



Resources & Tools III

- great task to come to a standard - its time
- can learn from initiatives such as JAXM etc
- let all parties who are working on this sit together in October

