

Technology – Broad View

Aspects that play a role when integrating archives leave the details of some core topics to the 2. day

Bernhard Neumair:

Thomas Soddemann:

Egon Verharen:

Peter Wittenburg:

Base Technologies

Middleware Concepts

Application Concepts

Interoperability Issues

Interoperability Aspects

- excellent paper by Bird and Simons
- many discussions about this topic during the last years
(Semantic Web, eScience)
- it will keep us busy despite all standardization efforts
D. Whalen: take all resources about EL you can get
- some general statements
(short since we know about the problems)
- more elaboration on metadata since
 - MD is important to get archives together
 - MD is not a central topic at this workshop

Interoperability Aspects

- technical encoding differences are still a problem for
 - texts (UNICODE not yet 100% coverage)
 - well-known to relevant boards
 - need better web-services for character conversion
 - media
 - shooting on a moving target due to technological development (storage/network + encoding/formats)
 - but conversion algorithms are available
 - in video area we miss same stability as in the audio area

Interoperability Aspects

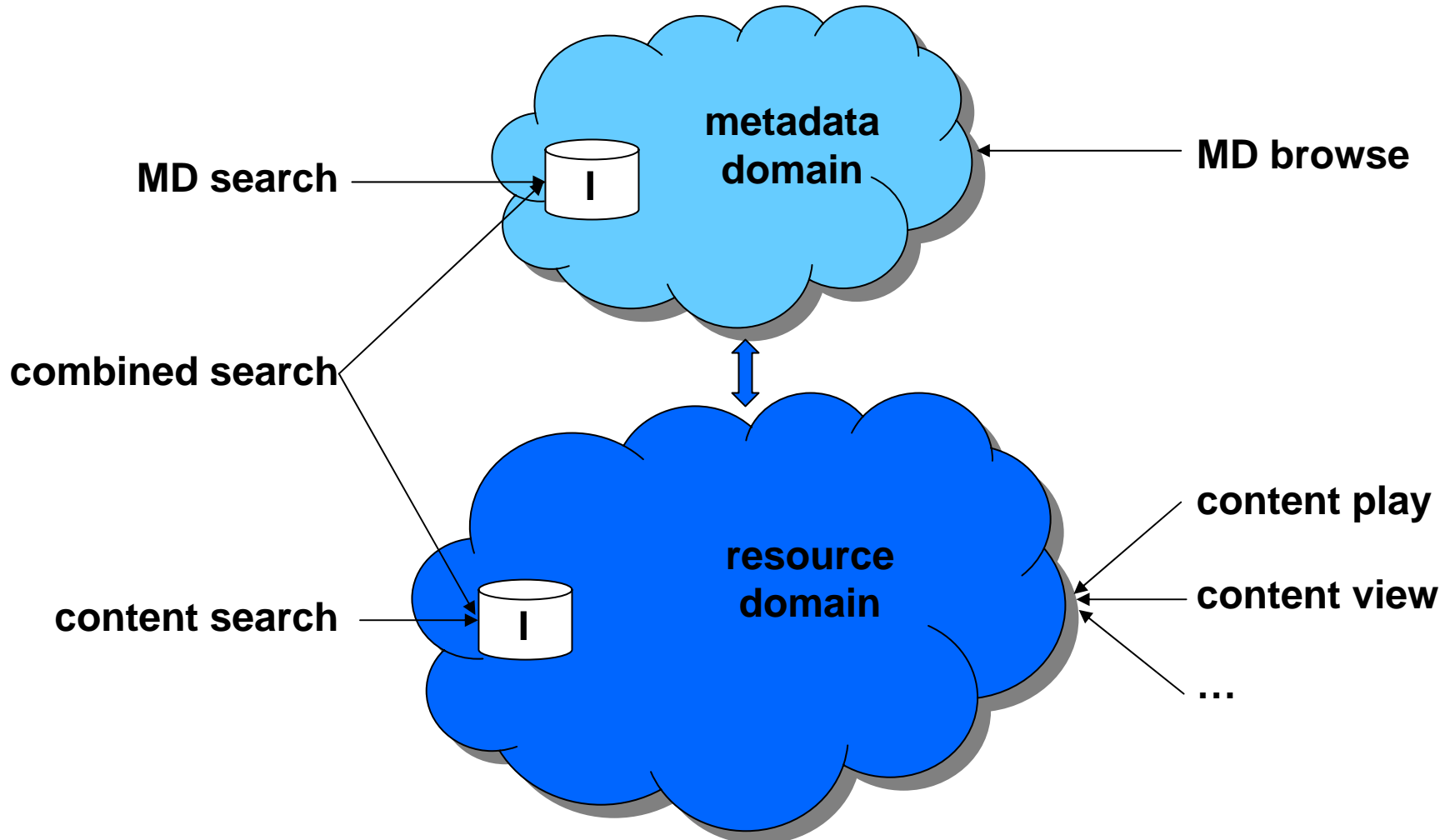
- differences in structure and structure description will remain
 - still many non-XML formats in use (legacy CHAT, WORD, ...)
 - still many resources are created without constraints (lots of errors)
 - XML does not solve the issue, however, it makes structure explicit
 - need agreements on general underlying models (AG, LAF, LMF, ...)
 - some conversions require interpretation
(what to do with inheritance and conditions)
 - debated in several meetings and initiatives (E-Meld, OLAC, ...)
 - need improved format conversion web-services

Interoperability Aspects

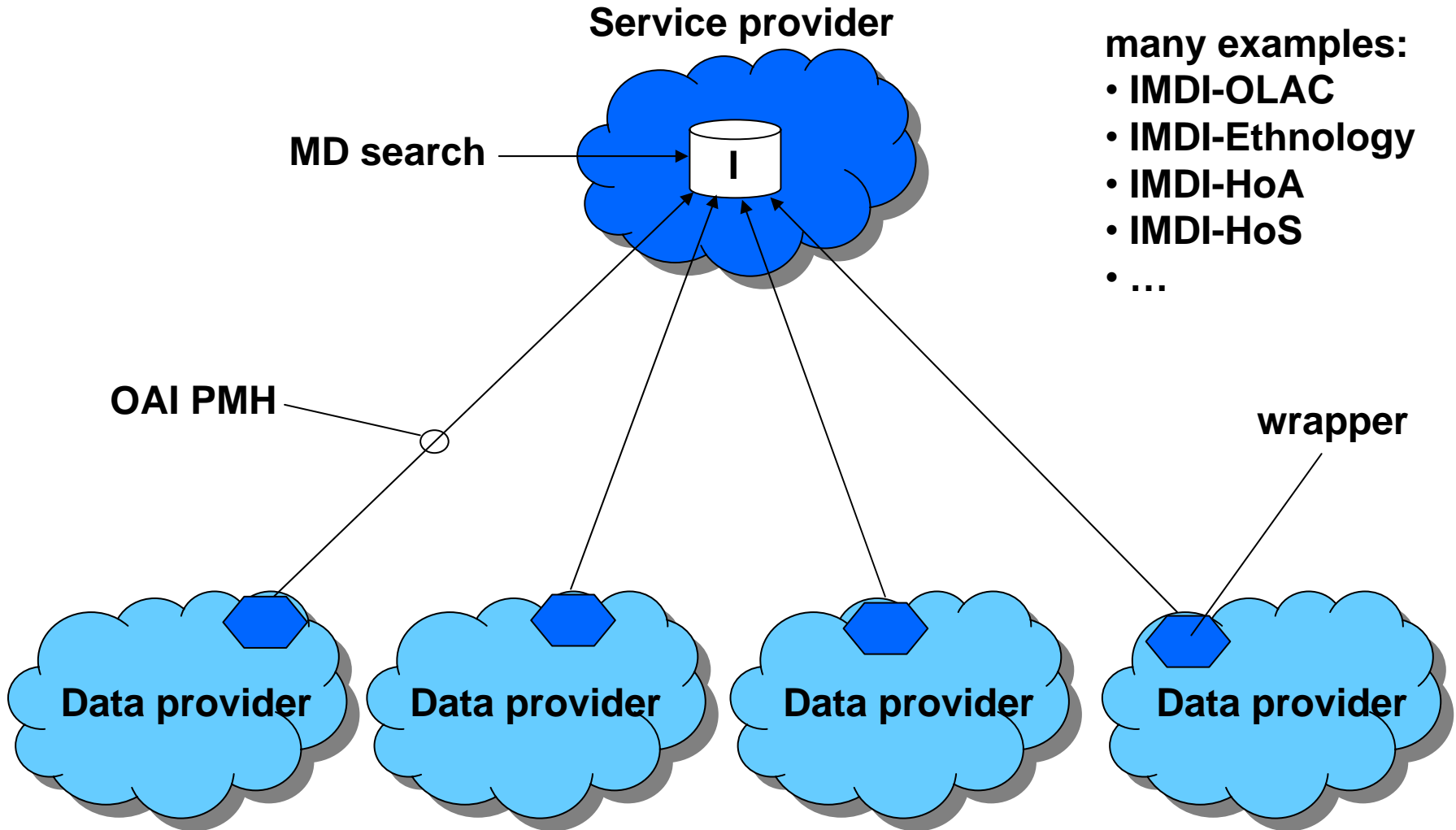
- linguistic encoding differences will remain
 - different theories and languages
 - ad hoc needs during fieldwork
 - also debated broadly during last years
 - formal frameworks are available (RDF, RDF-S, OWL)
- concrete activities to tackle these issues
 - GOLD ontology (E-Meld)
 - ISO TC37/SC4 Data Category Registry
 - IMDI-OLAC Gateway
 - ECHO DORA domain (10 repositories, 5 disciplines)
 - ...

Metadata – Classical View

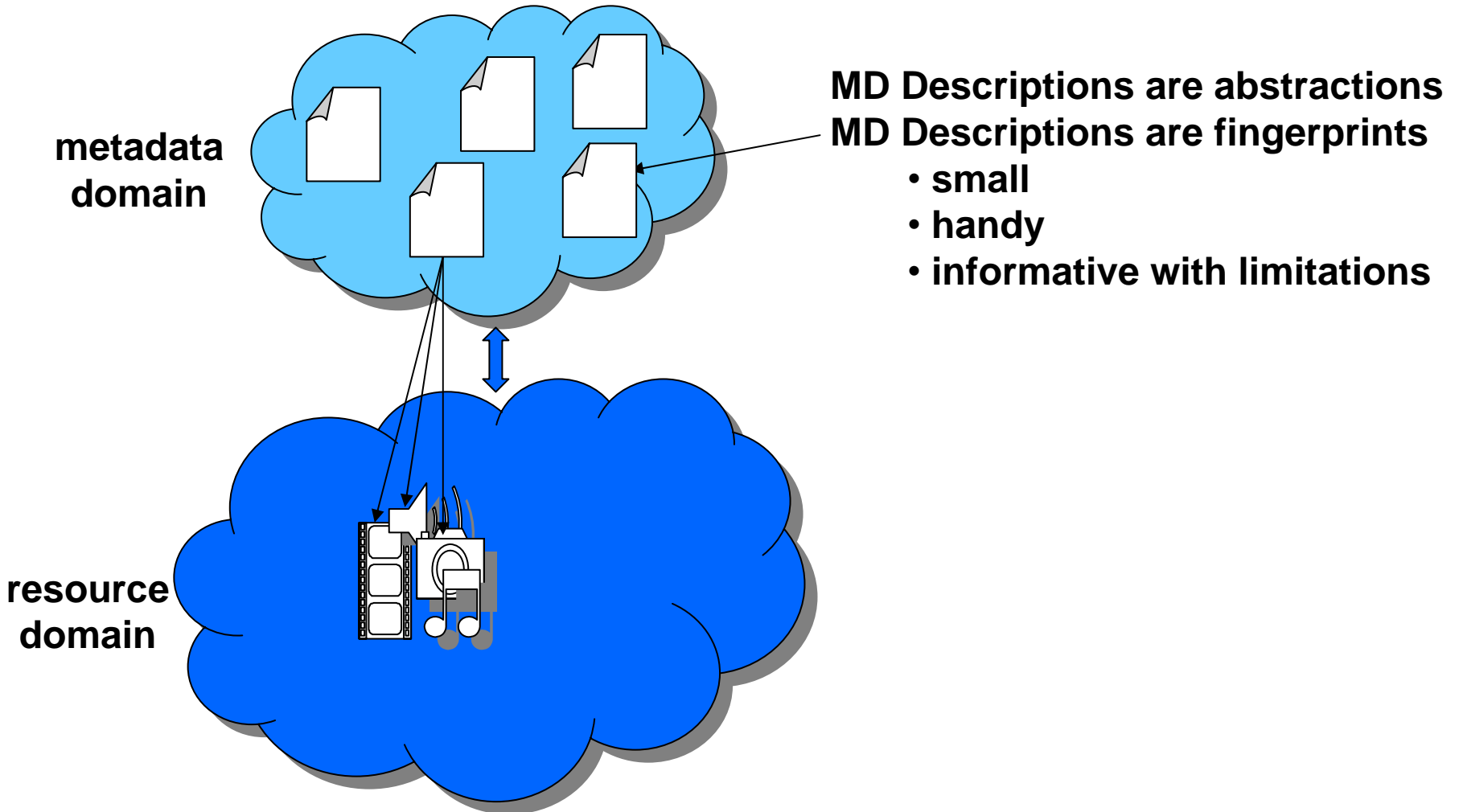
relevant for integrating archives



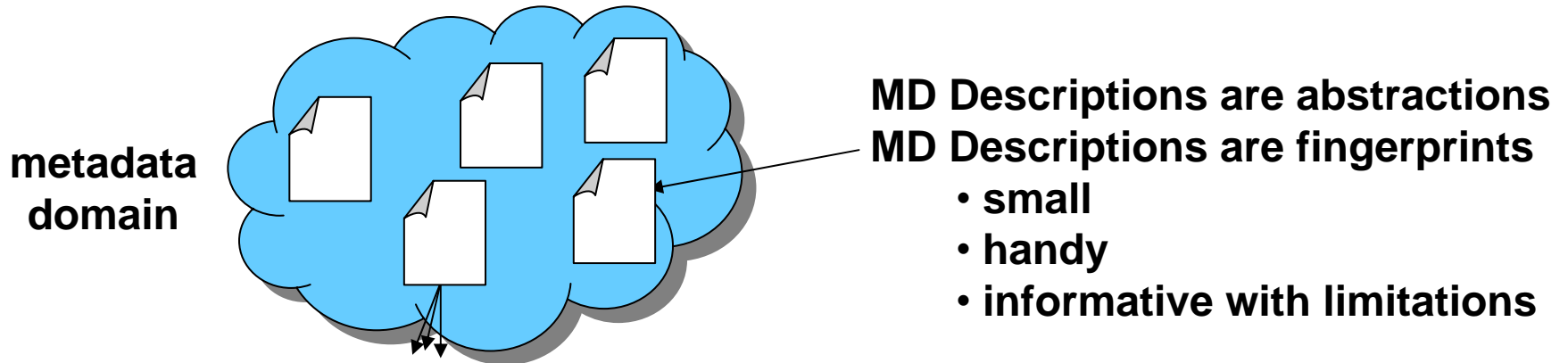
Metadata Classical View



Metadata Future View I



Metadata Future View I



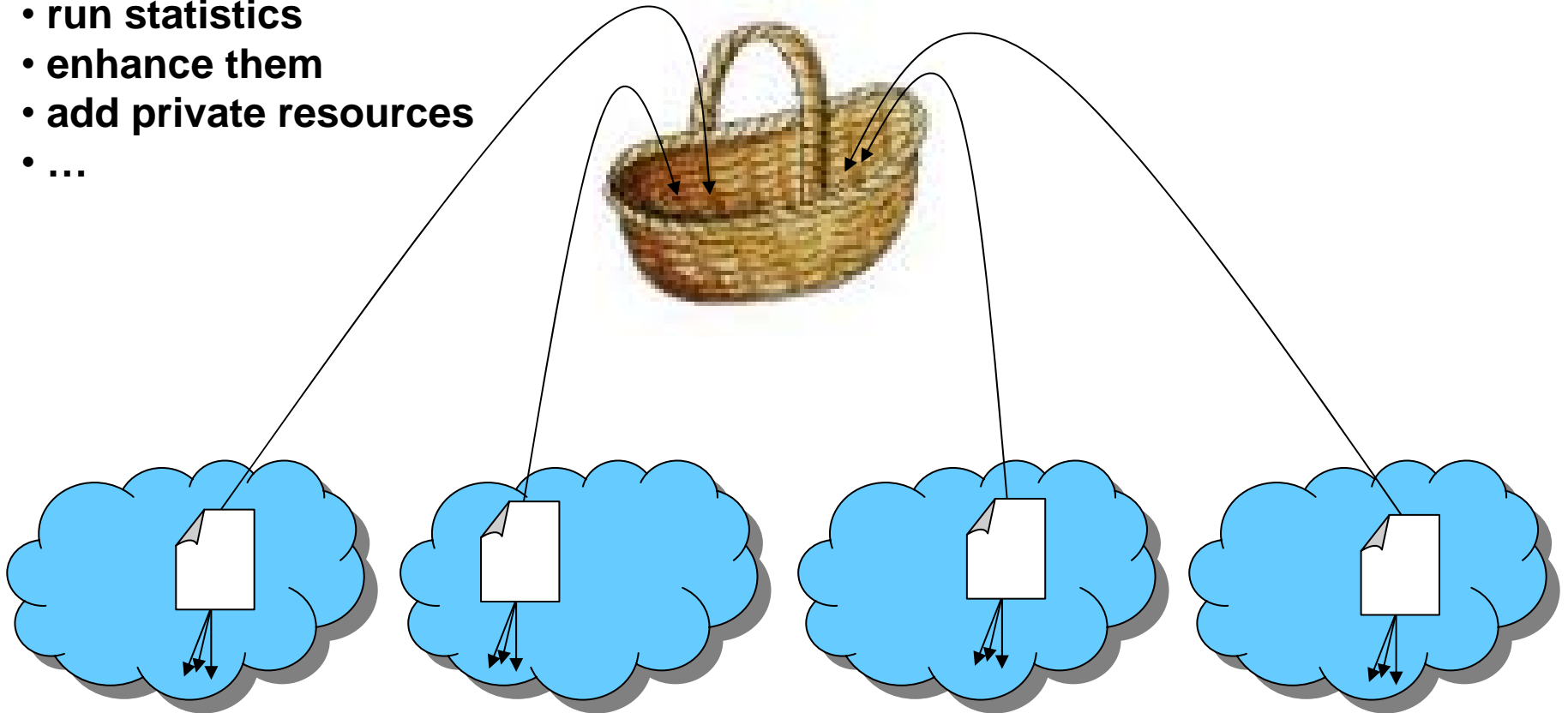
so – let's see what we can do with them
and what the requirements will be

Metadata Future View II

users can

- collect MD
- search on MD
- browse in MD
- run statistics
- enhance them
- add private resources
- ...

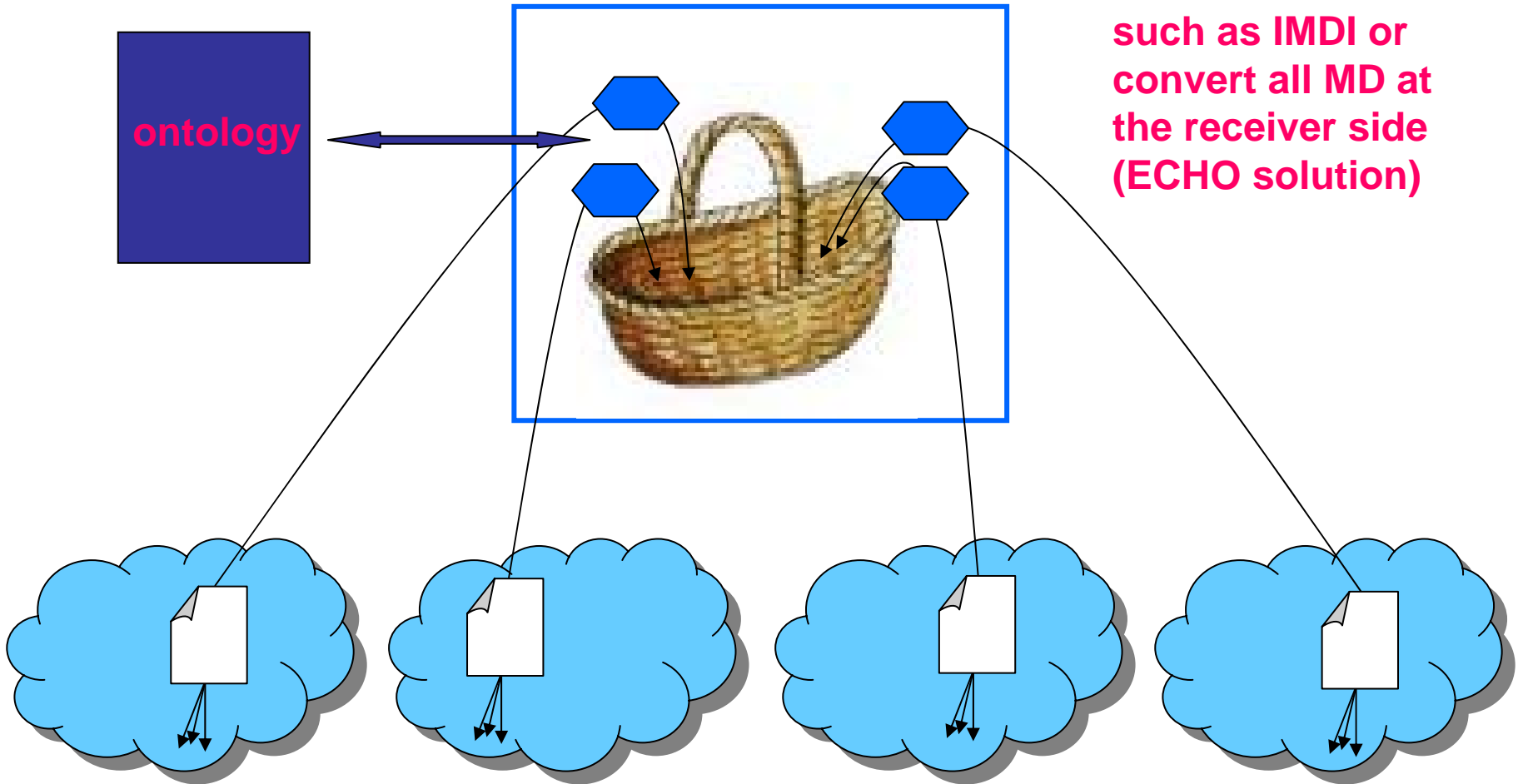
- basket has a new temporary personalized view on archival resources
- it's a private workspace
- maintain all references



Metadata Future View II

how can it work?

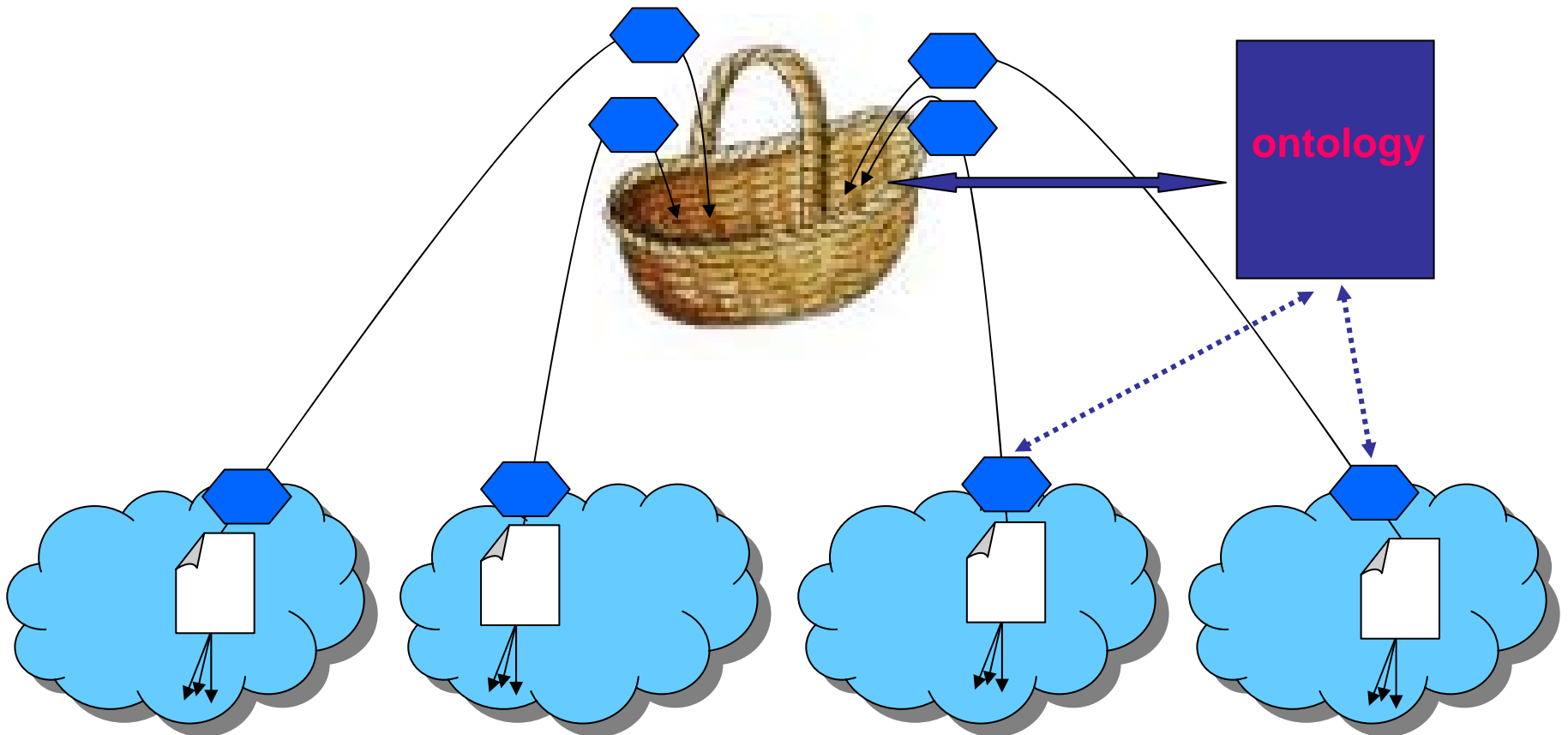
either remain within one domain such as IMDI or convert all MD at the receiver side (ECHO solution)



Metadata Future View III

- MD providers offer services structure is made explicit
- pick what you want and the way you want
- ontology still needed to do searches etc

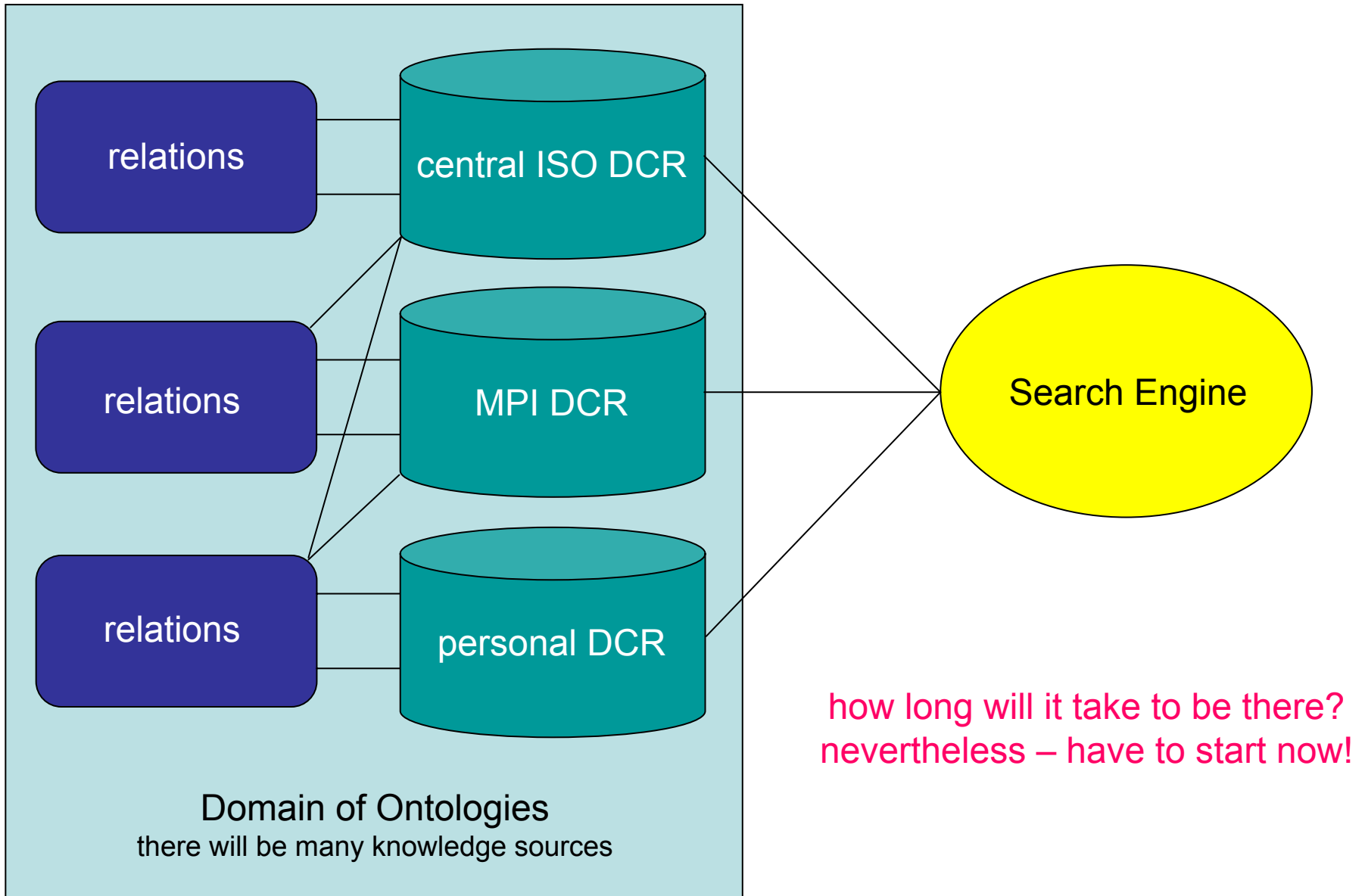
- who does what?
- will providers use ontologies?



Ontology Debate I

- **rich ontology vs. flat concept registry (incl. is_a relation)**
in case of flat registry: where to put all relations such as is_similar, has_a, ...
- **centralized ontologies vs. practical ontologies**
GOLD starts with SUMO
ISO DCR is central – agreement on personal DCRs

Ontology Debate II



Metadata Concepts

	MD	language	bundling	hierarchies	browsing	annotations
MPEG7	√		√	√	√	√
METS	√		√	√	√	
IMDI	√	√	√		√	
OLAC	√	√				
DC	√					

- **different container types are used (files, DB, CMS, ...)**
different shells/services necessary for exploration
- **“all” are schema based**

Metadata Future

- **did not speak about content integration problems are similar – text is not so much data**
- **what we want is clear or ?**
- **is it realistic?**
- **many questions to be answered**
- **it is an interesting time**