Integrating Data from The Perseus Project and Arachne using the CIDOC CRM

Robert Kummer, The Perseus Project
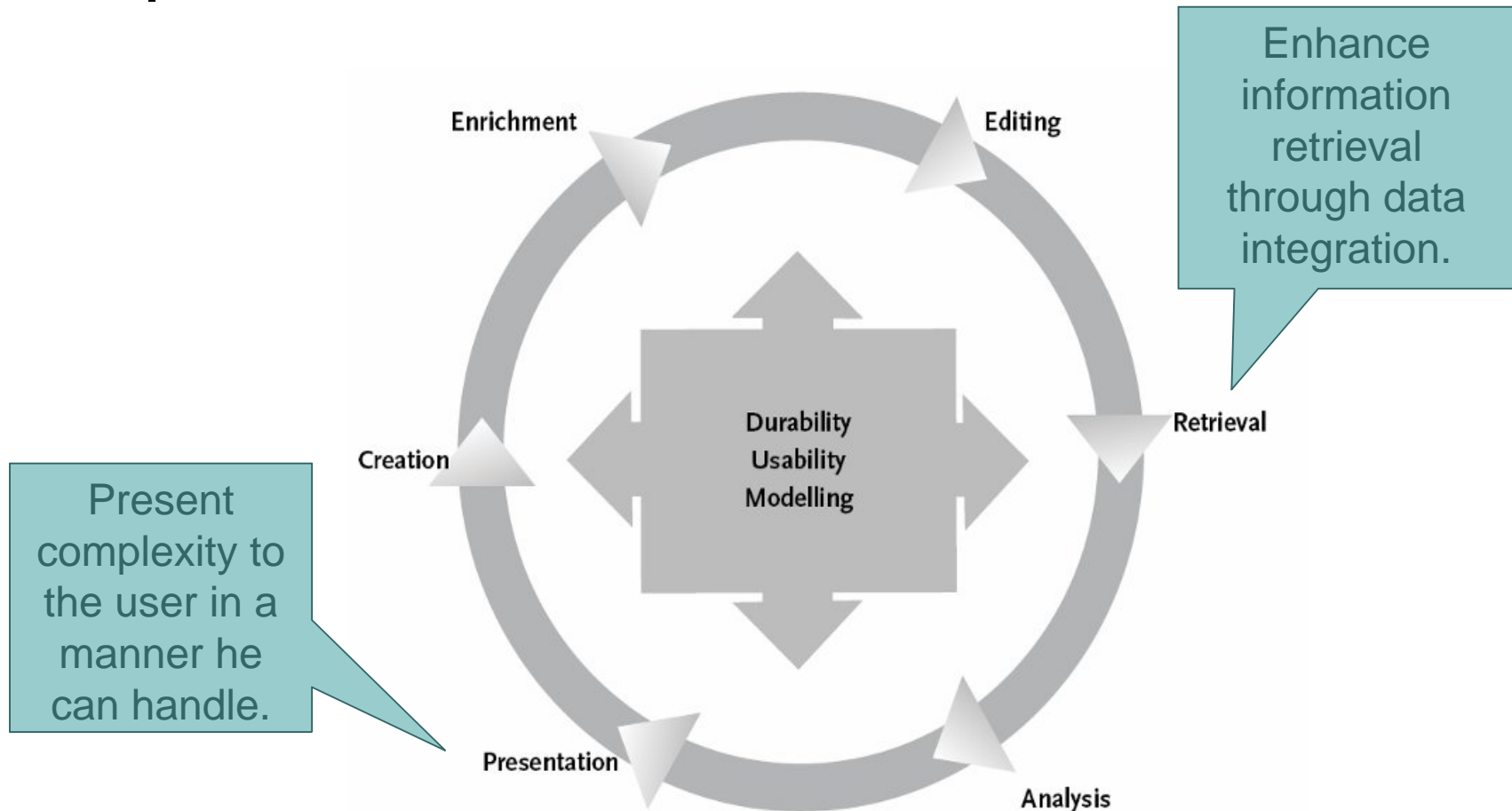
October 24, 2006

ICS-FORTH Workshop

Heraklion, Crete, Greece

Research Program

Enhance information retrieval through data integration.

Present complexity to the user in a manner he can handle.

Research Program II

- Enhance historical cultural research
  - Information integration
  - Provide linguistic assistance
  - Provide multilingualism
- Gain practice through implementation
  - What is already available?
  - What has to be built?
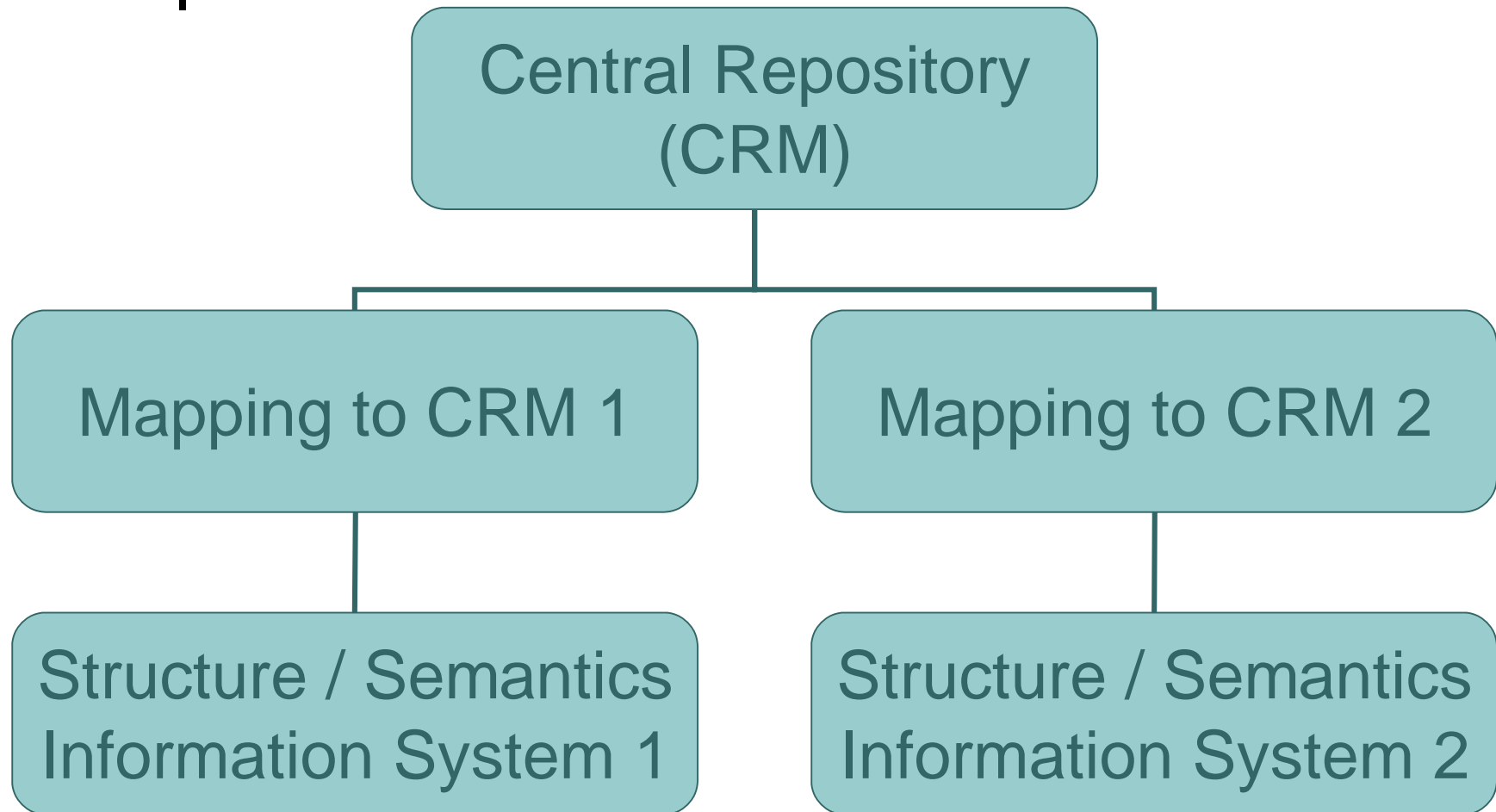 - Where do we encounter problems?

Focus on practice and implementation

- Due to lack of prior work
  - Focus on implementation
  - Focus on practice
- System architecture
  - Mapping agent for export of data
  - Central repository for storage and query

System Architecture

What has been achieved?

- Web Standards
  - XML/SQL for export
  - RDF/XML for representing CIDOC CRM
  - CIDOC CRM for representing domain knowledge
  - SPARQL for query formulation
- Software
  - Protégé
  - Jena API / Schemagen

Open questions

- Complexity
  - CRMish vs. CRMized
  - Theoretic discussion just evolving
  - Only few software available
  - Difficult estimations about time and effort
- Usability

Example: Ontobroker

Pictures taken from: Stefan Decker, Michael Erdmann, Dieter Fensel, Rudi Studer: How to Use Ontobroker, Karlsruhe.
URL: http://ksi.cpsc.ucalgary.ca/KAW/KAW98/decker/

Further Problems and Limitations

- Implementing the Mapping Agent
  - Implicit structure and semantics
  - Potential loss of information
  - Need for flexibility
- Implementing the Central Repository
  - Performance problems expected
  - Difficult graphical user interface

Implicit Structure and Semantics

- Database fields with internal structure
- External database objects
- Application layer and GUI
- Qualified relations

Arachne (Tables)	CIDOC CRM (Classes)	Comment
arachne -> object	E22.Man-Made_Object	
arachne -> refToPlace (M:N)	P54F.has_current_permanent_location	Relation is specified in Field arachne -> refToPlace -> typeOfPlaceRelation
arachne -> place	E53.Place	

Structure / Semantics (composition)

- Data model, how is factual knowledge stored

- Application logic, first layer of interpretation

- Graphical user interface, second layer of interpretation

- Users implicit knowledge, third layer of interpretation

Conclusion

- Use of the CIDOC CRM for information integration.
  - Provide a richer source of information.
  - Provide multilingual queries.
  - Keep the strength of each information system.
- Need to find ways to reduce complexity.
  - Identify important data for documentation and lookup.
  - Tune search tools to historian's needs (time, place, literature, contextualization).
  - Help the user to formulate queries and present data and structure in a manner laymen can handle.
  - Keep implementation time and effort within the bounds of possibility.
- Is a parameterized mapping agent a reachable goal?

Example "Traianscolumn"

Pictures taken from: http://www.arachne.uni-koeln.de/