



Waking from a Dogmatic Slumber -

A Different View on Knowledge Management for DL's

DELOS NoE Brainstorming Meeting

Martin Doerr

Center for Cultural Informatics
Institute of Computer Science
Foundation for Research and Technology - Hellas

Juan-les-Pins, France
December 5-6, 2005



Knowledge Management for DLs

Overview

***“There are no new research challenges in DL. There are only the ones from 30 years ago we still have not solved”
(anonymous, ECDL2005)***

Apologies: I’ll be deliberately provocative and possibly incomplete. Don’t take me too serious.

Overview:

- Use Cases
- Preconceptions and Solutions
- Understanding the Information Structure – Example ISO 21127
- Conclusions



Knowledge Management for DLs

Traditional Use Cases

Describing process knowledge (ontologies)

- ◆ ingest / cataloguing, preservation, certification, business models

Organization of information assets for access (ontologies)

- ◆ Metadata schemata
- ◆ Classification systems, subject headings (KOS)

Information for uniquely identified reference (KOS)

- ◆ citations, agents, places, objects, **events**, **periods** (facts)
- ◆ “concepts as facts” (species, types, theories, formula etc.) (ontologies)
- ◆ **unified KOS schemata** (Gazetteers, person lists, “related terms” in thesauri)

(challenges in red)



Knowledge Management for DLs

Advanced Use Cases

Information integration, use and reuse in the individual work environment

(see D.Soergel, D-Lib Magazine 2002, theme 2 & 3)

◆ Semantic Interoperability of metadata

- ontology-driven metadata schema integration
- mapping/merging of categorical data (reference information and classification concepts)
- matching factual data (“data cleaning”, reference information, identifier generation rules)

◆ Processing of user questions and user guidance

- search aid ontologies, ontology-supported natural language processing
- processing of “unknown” values and overlapping concepts in data and queries.

◆ Content integration for (re)use: The global network of knowledge

- retrieval of complementary information assets
- new linking schemes, matching factual data / identity over time
- knowledge extraction and integration



Knowledge Management for DLs

Grand Challenge

DLs should become integral parts of work environments as sources to find integrated knowledge and produce new knowledge.

But How ?

Knowledge management is the key. We distinguish:

1. Core ontologies for **schema semantics** such as metadata, content and KOS structures and processes. They are small and rich in relationships that **structure information** and relate content. They should be mutually compatible.
2. Ontologies as **categorical data** for reference and agreement as **objects of discourse**.
3. **Factual** background knowledge for reference and agreement as **objects of discourse**.

(see DELOS deliverable D5.3.1: "Semantic Interoperability in Digital Library Systems")



Knowledge Management for DLs

Preconceptions and Solutions

“Libraries should not depend on domain specific needs. Domains are too many and too diverse. Libraries need a generic approach.”

- ◆ This seduces us to only employ intuitive **top-down** approaches. As a result, when the fantasy is exhausted, research stops.
- ◆ **We need** deep knowledge engineering, generalizing in a **bottom-up** manner from real, specific cases to find the true generic structures across multiple domains. We need interdisciplinary work on **user-driven research scenarios**.

“Ontologies are huge, messy, idiosyncratic and domain dependent. Mapping is the only generic thing we can do”

- ◆ We are transfixed with ontologies used as categorical data, for which this statement is mainly true. We oversee the different character of ontologies describing **schema semantics**.
- ◆ These may pertain to **generic classes of discourse**, rather than to domains. They are the candidates to find generic relationships that integrated both factual and categorical knowledge in a way useful even for specific application. We need interdisciplinary work.



Knowledge Management for DLs

Preconceptions and Solutions

“Queries are mainly about classes. The main challenge of information integration is the integration of (ontologies of) classes.”

- ◆ We believe this is **not sufficiently** supported by empirical studies. The question: “do we talk about the same thing?” pertains to **universals and particulars**.
- ◆ **We need** to systematically **analyze the workflow** of research work and the **original research questions** in each phase. We need to provide access by **factual relationships** (Amit Sheth), such as “georeferencing”. We need interdisciplinary work for generalizations.

“Manual work is not scalable or affordable. Only fully automated methods have a chance”

- ◆ This seduces us to **discard the quality** of manual, intellectual decisions. Yet billions of people produce content manually. Wikipedia demonstrates, that the above is not true.
- ◆ **We need** to design the interactive processes and the awarding to massively involve **Virtual Communities / Organisations** in cataloguing, data cleaning and ontology development. We need **semiautomatic** algorithms. We need interdisciplinary work.



Knowledge Management for DLs

Do we talk about the same thing?

“We need more reasoning!”

- ◆ This is true. But **what sort of** reasoning? And before **any reasoning** can be done, data must be connected, in a **“global network of knowledge”**. We must first clarify: *Do we talk about the same thing?*

- ◆ We need:
 1. massive analysis of **true research questions** to find the correct relationships and the correct reasoning forms.
 2. core ontologies with the **revelant relationships**.
 3. **knowledge extraction** to populate the networks.
 4. theories of negotiating **(diachronic) identity** of referred items across contexts.
 5. algorithms for global, massive, **semiautomatic data cleaning** across contexts and to restore referential integrity in order to create, maintain and improve global networks of knowledge.
 6. large **distributed KOS** with generic architectures to provide persistency to the data cleaning process.

- ◆ And **only then** we can do advanced reasoning and intelligent query processing ...



Knowledge Management for DLs

Example: the core ontology ISO21127

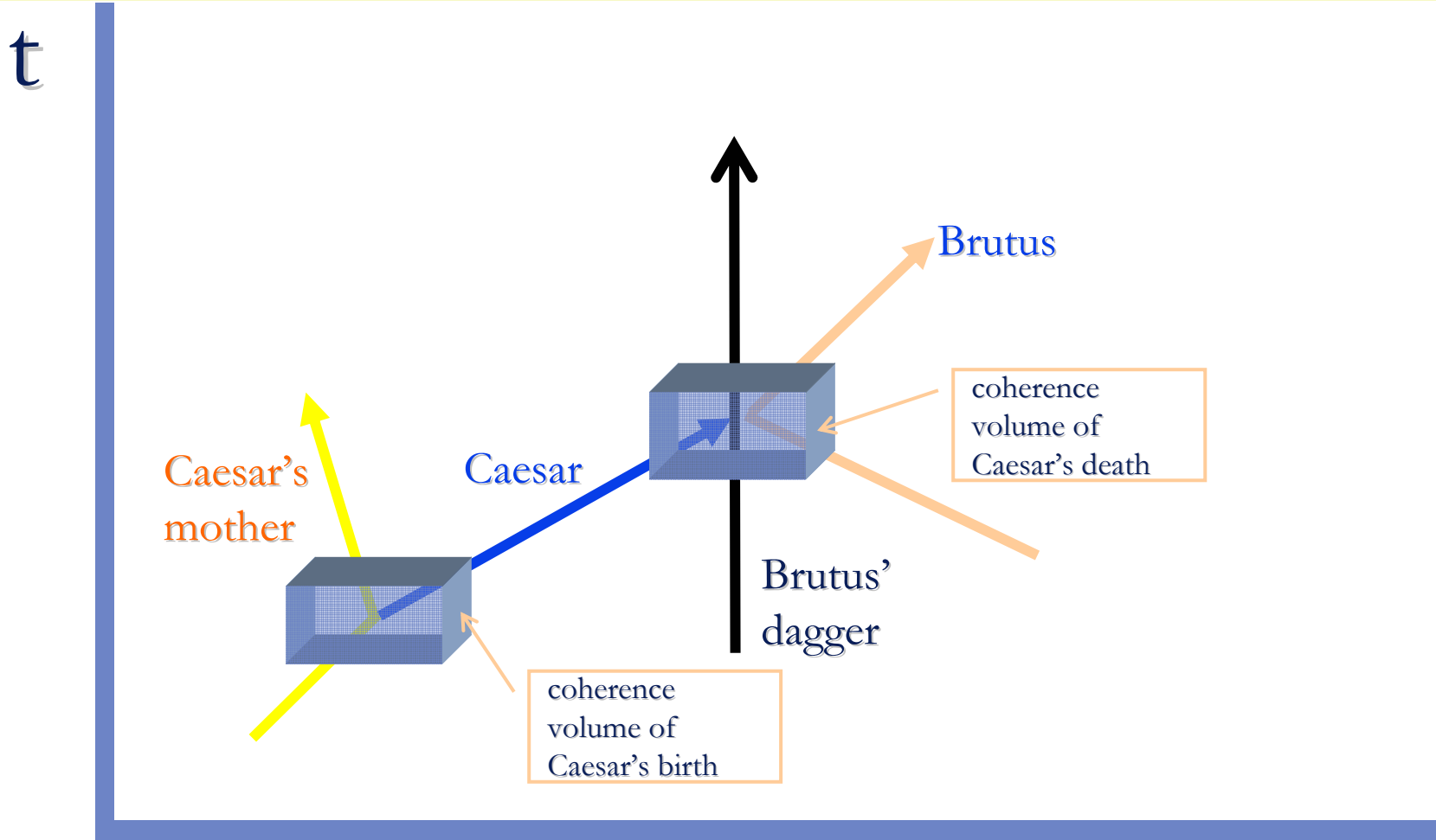
The CIDOC Conceptual Reference Model (ISO/FDIS 21127)

- ◆ is a **core ontology** describing the underlying semantics of data schemata and structures from all museum disciplines and archives. Now being merged with **IFLA FRBR** concepts (*DELOS WP5*).
- ◆ It is result of long-term **interdisciplinary work** and agreement.
- ◆ In essence, it is a **generic model** of recording of “what has happened” in human scale, i.e. a class of discourse.
- ◆ It can generate huge, meaningful **networks of knowledge** by a simple abstraction: history as meetings of people, things and information.
- ◆ **It bears surprise**: more effective metadata structures, and linking schemes



Knowledge Management for DLs

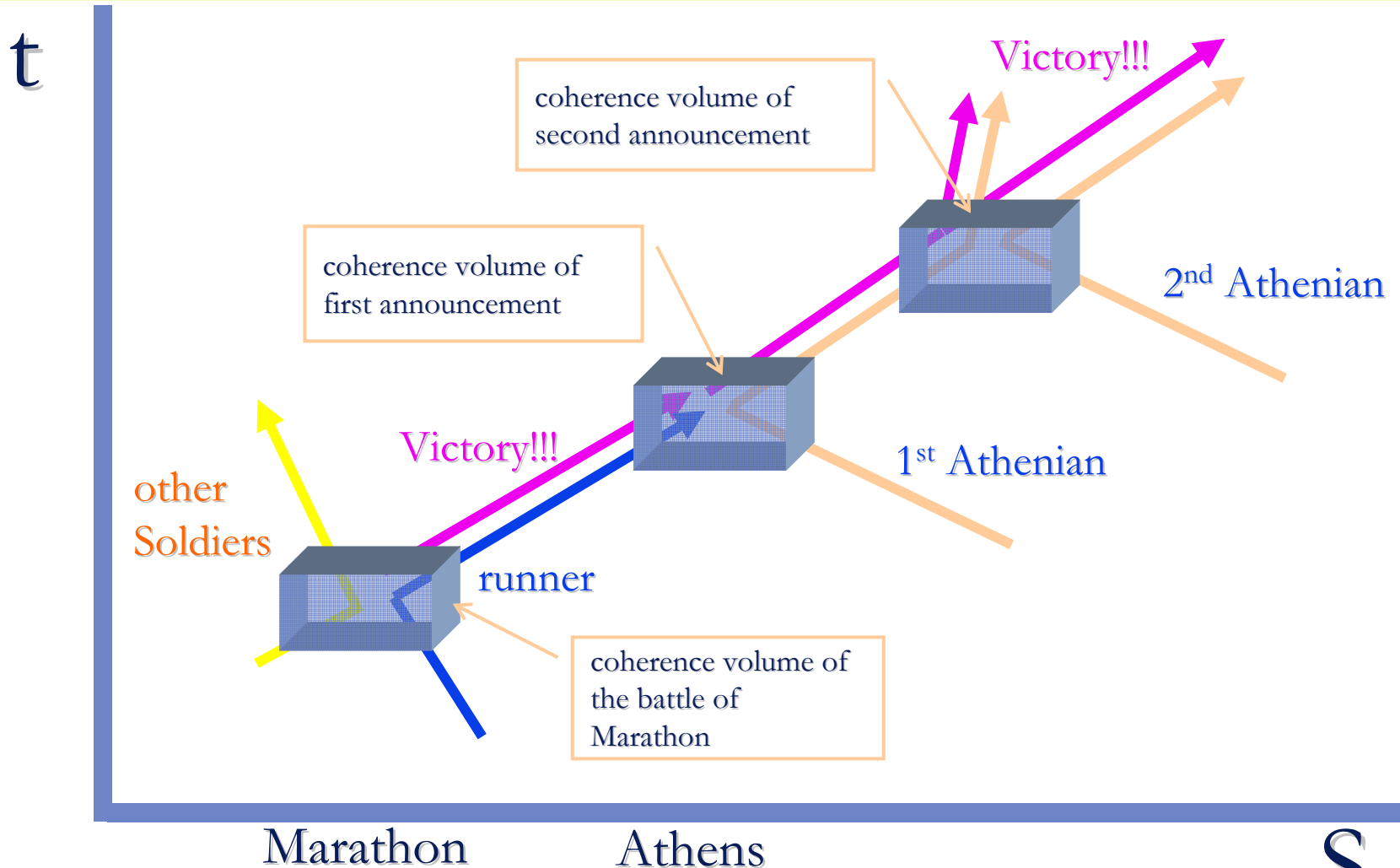
Example: History as Meetings of People..





Knowledge Management for DLs

Example: ...Things and Information





Knowledge Management for DLs

Example: Meetings and Metadata

Type:	Text
Title:	Protocol of Proceedings of Crimea Conference
Title.Subtitle:	II. Declaration of Liberated Europe
Date:	February 11, 1945.
Creator:	The Premier of the Union of Soviet Socialist Republics The Prime Minister of the United Kingdom The President of the United States of America
Publisher:	State Department
Subject:	Postwar division of Europe and Japan

Metadata

Documents

About...

“The following declaration has been approved:
The Premier of the Union of Soviet Socialist Republics,
the Prime Minister of the United Kingdom and the President
of the United States of America have consulted with each
other in the common interests of the people of their countries
and those of liberated Europe. They jointly declare their mutual
agreement to concert...
....and to ensure that Germany will never again be able to
disturb the peace of the world..... “



Knowledge Management for DLs

Example: Meetings and Metadata

Type:	Image
Title:	Allied Leaders at Yalta
Date:	1945
Publisher:	United Press International (UPI)
Source:	The Bettmann Archive
Copyright:	Corbis
References:	Churchill, Roosevelt, Stalin

Metadata



About...

Photos, Persons

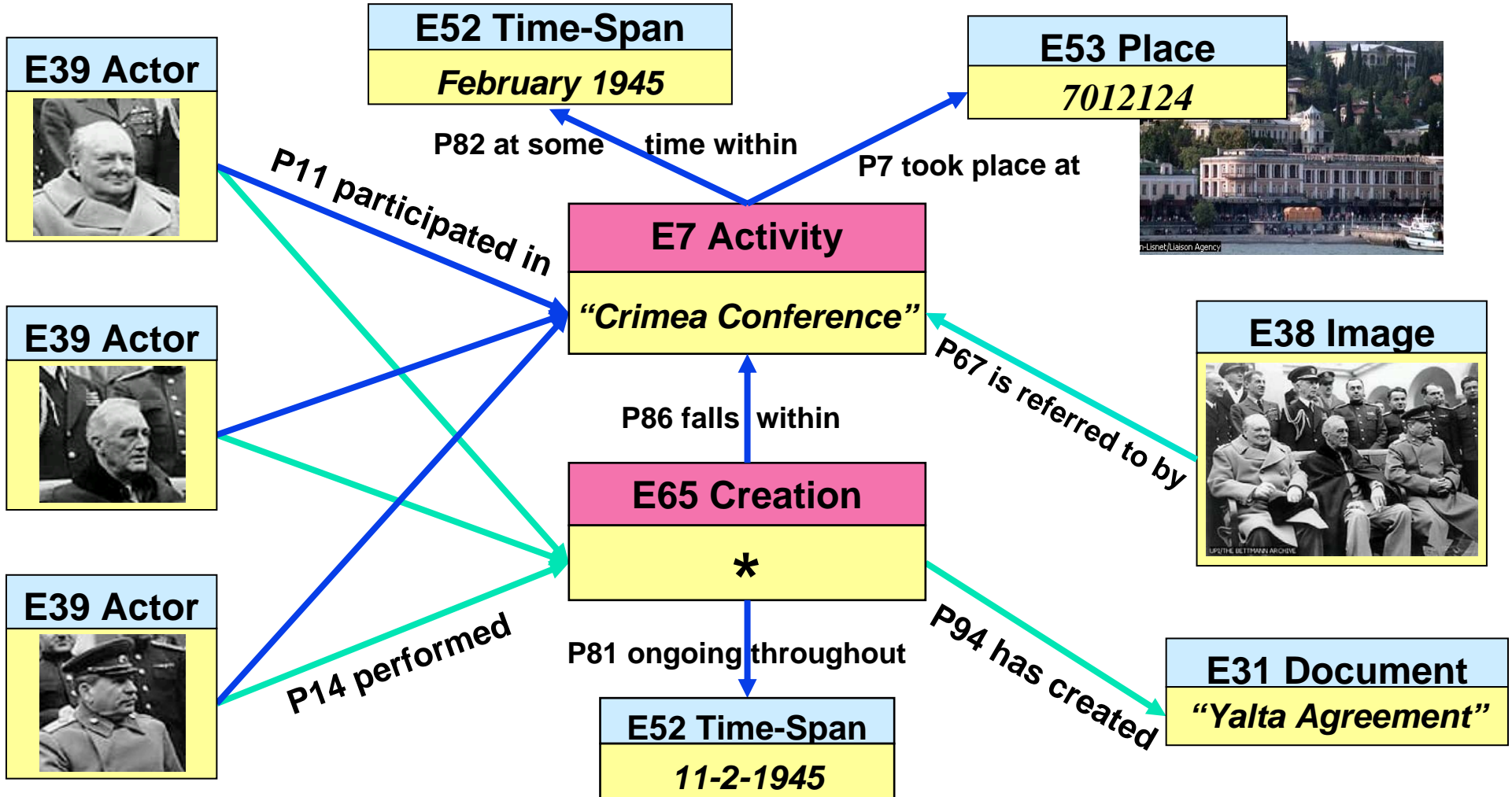


UPI/THE BETTMANN ARCHIVE



Knowledge Management for DLs

Example: The ISO21127 Solution

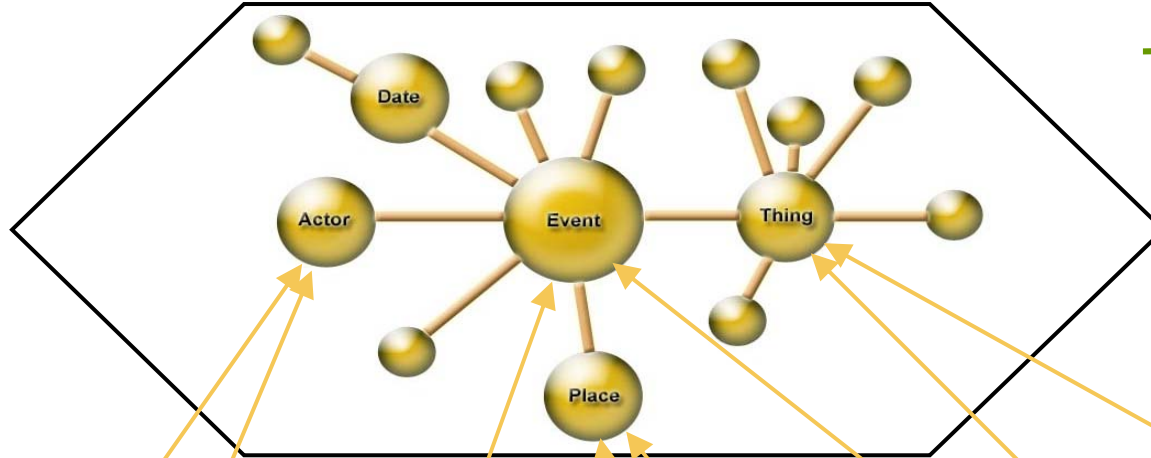




Knowledge Management for DLs

Example: Are Hyperlinks Wrong?

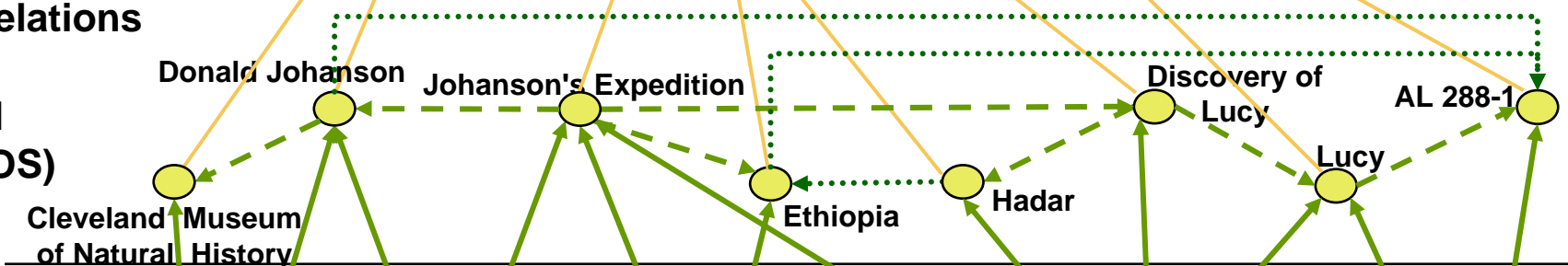
CIDOC CRM
Core Ontology



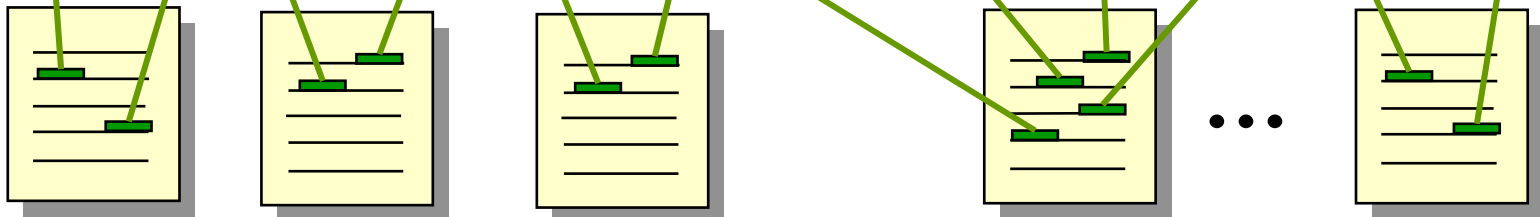
- Linking documents by co-reference
- - - → Primary link corresponding to one document
- ⋯ → Deductions
- Instance of

Integration by
Factual Relations

real world
nodes (KOS)



Documents in
Digital Libraries





Knowledge Management for DLs

Conclusions

If we rethink old positions, we will find surprising new answers to

“..an information model for digital libraries that intentionally moves 'beyond search and access', without ignoring these functions, and facilitates the creation of collaborative and contextual knowledge environments.”

(C.Lagoze, D-Lib Magazine 2005)

But:

- ◆ We need a **massive investment in understanding** and generalizing the intellectual processes and original **research questions**.
- ◆ We have to do basic research in **dynamic collaborative knowledge organization forms**, formal processes and algorithms that **converge** to higher stages of knowledge.
- ◆ The large networks of integrated knowledge to come will need continuous maintenance with **new, specific social organisation forms** and GRID-like resource access, and they may look very different from our current systems...

(This is again a 30 years old dream, are we closer now?)



Knowledge Management for DLs

Conclusions

If we rethink old positions, we will find surprising new answers to

“..an information model for digital libraries that intentionally moves 'beyond search and access', without ignoring these functions, and facilitates the creation of collaborative and contextual knowledge environments.”

(C.Lagoze, D-Lib Magazine 2005)

- ◆ The step to integrate digital libraries in working environments will need a massive investment in knowledge engineering and generalization of the respective intellectual processes in **all key disciplines**.
- ◆ It is feasible to envisage large networks of integrated knowledge. They will however need **continuous maintenance** with new, specific organisation forms and GRID-like resource access.
- ◆ On these premises, we can identify new, specific research tasks and strategies to integrate the results into **functional** working infrastructures.
- ◆ And find