

International standards, national traditions: Vocabulary control in archival science

Jonathan Furner
UCLA

Outline

- the InterPARES project
 - objectives and methods
- its terminological database
 - opportunities and challenges

10/17/06

SAA Annual Meeting,
Boston, 2004

2

Some basic definitions

- **record**: a document that (a) participates in or results from the activity of an individual or organization, and (b) constitutes the primary source of knowledge about that activity
 - increasingly likely to be generated in electronic form
- **preservation**: the physical and intellectual protection of records through time
- **obstacles** to long-term preservation:
 - obsolescence, incompatibility of software, hardware
 - fragility of storage media
 - hybridity of record formats
 - **ease of manipulation**

10/17/06

SAA Annual Meeting,
Boston, 2004

3

Some basic definitions, cont'd

- **authentic record**: a record that can be proved to be what it purports to be
 - free of tampering or corruption
 - trustworthy as a record
 - authenticity conferred by
 - mode, form, state of transmission as drafts, originals, copies
 - manner of maintenance, preservation, custody
- **authentication**: a means of verifying the authenticity of a record at time *t*
 - e.g., the insertion of an element or statement in a record
 - governed by rules established by legislation
- **problem**: how to evaluate the authenticity of **electronic** records?

10/17/06

SAA Annual Meeting,
Boston, 2004

4

The InterPARES project

- International Research on Permanent Authentic Records in Electronic Systems
- **aim**: to formulate principles / criteria ...
- ... for development of policies / strategies / standards ...
- ... for the long-term preservation of authentic electronic records

10/17/06

SAA Annual Meeting,
Boston, 2004

5

InterPARES, cont'd

- **international** ...
 - Asia: China, Japan, Singapore
 - Australia
 - Canada
 - Europe: Belgium, France, Ireland, Italy, Netherlands, Portugal, Sweden, UK
 - USA
- ... and multidisciplinary
 - archivists, computer scientists, librarians
 - artists, humanists, natural scientists, social scientists

10/17/06

SAA Annual Meeting,
Boston, 2004

6

InterPARES 1 (1999-2002)

- conceptual **requirements** for authenticity of various types of electronic records
- **methods** of / criteria for
 - selection / appraisal of electronic records
 - preservation of ability to reproduce electronic records
- intellectual framework for development of **policy**
- **instruments** for analysis/assessment of existing records
- **glossary**

10/17/06

SAA Annual Meeting,
Boston, 2004

7

InterPARES 2 (2002-)

- **methods** of ensuring reliability, accuracy, authenticity
- **policies**/strategies for overcoming media fragility, technological obsolescence
- focus on **complex records** typically created in experiential, interactive, dynamic systems ...
- ... in course of artistic, scientific, e-government activities
- **understanding** of intellectual property, data privacy issues
- **thesaurus**

10/17/06

SAA Annual Meeting,
Boston, 2004

8

Research methods

- “traditional” methods
 - surveys and case studies
 - grounded theory
 - diplomatic analysis
 - aim: to assess authenticity of records of unverified provenance, independently of context

10/17/06

SAA Annual Meeting,
Boston, 2004

9

Research methods, cont'd

- “nontraditional” methods
 - modeling of **activities**
 - hierarchical decomposition
 - identification of constraints, motivations, requirements, products (i.e., entities playing different roles)
 - modeling of **entities**
 - identification of **attributes, relationships**
 - identification of entity types: e.g., agents, events, objects ...
 - **terminological** analysis
 - consistent, rigorous definition of all terms used to denote activities, entities, attributes, etc.

10/17/06

SAA Annual Meeting,
Boston, 2004

10

Modeling example

- **activity**: Manage Chain of Preservation
 - **constraints**: Juridical System; State of Technology; Preserver's Mission; Archival Principles ...
 - **motivations**: Information About ... Documents, Documents' Context, Technology, Records Creator ...
 - **requirements**: Creator; Records Manager; Preserver ...
 - **products**: Sent Records; Preserved Records; Documentation About Destroyed Records ...

10/17/06

SAA Annual Meeting,
Boston, 2004

11

Alternative frameworks

- a **document life cycle** model: top-level activity = “Manage chain of preservation”
 - creation → recordkeeping → selection and preservation
- a **records continuum** model
- a **unified** model: top-level activity = “Manage records management life cycle”
 - manage life cycle framework → carry out business activity → manage records

10/17/06

SAA Annual Meeting,
Boston, 2004

12

Potential disagreements

- definitions of basic terms: e.g., "record," "archives"
 - esp. in electronic environment, where records are defined in terms of the **context** of creation
- role of archivist
 - keeper (cf. Jenkinson) vs. selector (cf. Schellenberg)
- criteria for selection and appraisal
 - function (cf. J) vs. use (cf. S)
 - context (inc. provenance) vs. form vs. content

10/17/06

SAA Annual Meeting,
Boston, 2004

13

Purpose of terminological tools

- to improve the effectiveness of IP-related **communication** and **dissemination**
- by helping **authors**, **readers**, and **intermediaries** to select and use terms **consistently**
- thus increasing the likelihood ...
 - that authors will use terms in the manner expected by their readers: effective **authorship**
 - that readers will find documents of the kind they want: effective **retrieval**
 - that readers will interpret documents in the manner intended by their authors: effective **understanding**

10/17/06

SAA Annual Meeting,
Boston, 2004

14

Some general challenges

- potential **conflict** of purpose: is partly **descriptive**, partly **prescriptive**
 - founded on current usage ...
 - ... but specifying preferred future usage
- **specificity** of (often multiple) senses of **technical** terms
- **domain**-specific variation
- **region**-specific variation
 - requirements for authenticity are not established by legislation, but by cultural tradition
 - thus, contextualization of project findings in national, multinational, international contexts is essential

10/17/06

SAA Annual Meeting,
Boston, 2004

15

Some general challenges, cont'd

- multiplicity of (sometimes contradictory) vocabulary-construction **standards**
 - *ISO 2788: Guidelines for the establishment and development of monolingual thesauri*, 2nd ed. Geneva: International Organization for Standardization, 1986.
 - *ANSI/NISO Z39.19: Guidelines for the construction, format and management of monolingual thesauri*. Bethesda, MD: National Information Standards Organization, 1993.
- **facet** analysis

10/17/06

SAA Annual Meeting,
Boston, 2004

16

Four terminological tools

- Register
- Dictionary
- Glossary
 - Access
- Thesaurus
 - TCS-8
- four views of one database

10/17/06

SAA Annual Meeting,
Boston, 2004

17

Core data elements

- **qualified term** [register]
 - term
 - part-of-speech
 - facet
 - domain
- **definition** [dictionary]
 - sense
 - source
- **approval** status (as IP-preferred sense) [glossary]
- **contextualization** [thesaurus]
 - broader term
 - related term

10/17/06

SAA Annual Meeting,
Boston, 2004

18

Terms, qualifiers, definitions

- a **qualified term** is made up of ...
 - one **term**: e.g., "record"
 - three **qualifiers**:
 - a **part of speech**: e.g., "noun"
 - a **facet**: e.g., "object"
 - a **domain**: e.g., "archival science"
- a single term (**homonym**) may have multiple **senses** / meanings, and ...
- ... different terms (**synonyms**) may have the same sense
- aims:
 - disambiguate homonyms (through qualifiers)
 - collocate synonyms (through links between nonpreferred and preferred terms)

10/17/06

SAA Annual Meeting,
Boston, 2004

19

Auditing and reporting

- **audit trail**
 - **suggester**: name; date
 - last **modifier**: name; date
 - **accepter**: name; date
 - **rejecter**: name; date
 - all displayed in Register view
- **weekly reports** (to terminology team; automatically generated)
 - suggestions and modifications (for acceptance/rejection/modification)

10/17/06

SAA Annual Meeting,
Boston, 2004

20

Dictionary records: Example

Term	PoS	Facet	Domain	Sense	Source
record	n.	object	arch. sci.	A document created or received and maintained by an agency, organization, or individual in the pursuance of legal obligations or in the transaction of business	IP
record	n.	object	arch. sci.	Document created by a physical or juridical person in the course of practical activity	UBC
record	n.	object	comp. sci.	A grouping of interrelated data elements forming the basic unit of a file	SAA

10/17/06

SAA Annual Meeting,
Boston, 2004

21

Dictionary

- **vocabulary**:
 - all defined terms
 - one or more definitions per qualified term
 - one record per definition
- **arrangement**:
 - alphabetical
- **data**:
 - senses + sources

10/17/06

SAA Annual Meeting,
Boston, 2004

22

Glossary records: Example

Term	PoS	Facet	Domain	Sense	Approved?
record	n.	object	arch. sci.	A document created or received and maintained by an agency, organization, or individual in the pursuance of legal obligations or in the transaction of business	Yes
record	n.	object	comp. sci.	A grouping of interrelated data elements forming the basic unit of a file	Yes

10/17/06

SAA Annual Meeting,
Boston, 2004

23

Glossary

- **vocabulary**:
 - IP- "approved" [preferred] terms
 - one definition per qualified term
 - one record per qualified term
- **arrangement**:
 - alphabetical
- **data**:
 - IP- "approved" [preferred] sense

10/17/06

SAA Annual Meeting,
Boston, 2004

24

Thesaurus records: Example

Term	Facet	Subfacet	Link type	Linked term
record	object	[none]	NTG	document
record	object	[none]	RT	archives
active record	object	records by level of activity	NTG	record
reliable record	object	records by level of reliability	NTG	record

10/17/06

SAA Annual Meeting,
Boston, 2004

25

Thesaurus

- **vocabulary:**
 - all defined terms (cf. Dictionary)
- **arrangement:**
 - faceted [not, e.g., disciplinary] and hierarchical
- **data:**
 - links among paradigmatically-related terms
 - hierarchical
 - genus–species [as NTGs]
 - whole--part [as RTs]
 - [class--instance]
 - from IP-nonpreferred to IP-preferred

10/17/06

SAA Annual Meeting,
Boston, 2004

26

Current facets

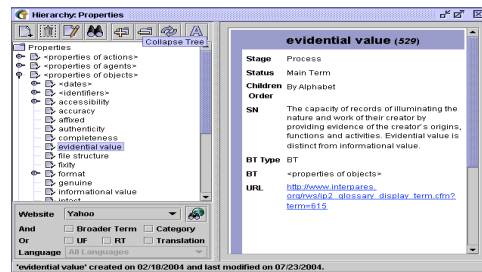
- **objects** [inc. sets and parts of objects]
 - e.g., “fonds”; “active record”
- **actions** [i.e., activities, processes]
 - e.g., “authentication”; “emulation”
- **agents**
 - e.g., “juridical person”; “creator”
- **properties** [i.e., attributes, characteristics, of objects, actions, agents]
 - e.g., “authenticity”; “date of receipt”
- **disciplines?; concepts?; events?; systems?; principles?**

10/17/06

SAA Annual Meeting,
Boston, 2004

27

TCS-8 example



10/17/06

SAA Annual Meeting,
Boston, 2004

28

Some specific challenges

- selection of additional vocabulary
 - which sources should be consulted?
 - what procedures and criteria should be used?
 - should compound terms be selected? (if so, how?)
- level of pre-coordination
- assignment of terms to facets
 - is the current facet-set useful and/or appropriate?
- categorization of terms within facets
 - within each facet, what set of categories is useful and/or appropriate?

10/17/06

SAA Annual Meeting,
Boston, 2004

29

Some specific challenges, cont'd

- reconciliation of conflicting recommendations, in the various standards, as to form of terms
- reconciliation of conflicting recommendations, by IP stakeholders, as to definition of terms
- workflow
- **evaluation**
 - how are tools to be evaluated w.r.t.
 - complying with international standards?
 - meeting users' requirements?

10/17/06

SAA Annual Meeting,
Boston, 2004

30

Conclusion

- the process of deciding what is worth remembering and what is not is **political** (cf. Cook, 1997)
- so is the process of prescribing (“standardizing”) the meanings of terms
 - no lexicographer is neutral or impartial
- tests: consistency, simplicity, utility

10/17/06

SAA Annual Meeting,
Boston, 2004

31

Terminology team

- lexicographer: Naomi Cull (Toronto)
- database administrator: Jean-Pascal Morghese (UBC)
- graduate student researcher: Nadav Rouche (UCLA)
- Barbara Craig
- Luciana Duranti
- Phil Eppard
- Jonathan Furner: jfurner@ucla.edu
- Ian Lancashire
- Richard Pearce-Moses
- John Roeder

10/17/06

SAA Annual Meeting,
Boston, 2004

32