



What is metadata and what is it good for?

Peter K. Austin

Endangered Languages Academic Programme
SOAS, University of London

Tutorial
LSA Annual Meeting, Pittsburgh
7 January 2011

www.hrelp.org



Metadata

- metadata is *data about data*
 - for **identification, management, retrieval** of data
 - provides the **context** and **understanding** of that data
- carries those understandings into the **future**, and to **others** (and hence is important for archiving and preservation)
- reflects **knowledge** and **practices** of data providers
- **defines** and **constrains** audiences and usages for the data



Metadata

- recommendations for creating metadata for language documentation have been primarily influenced by library concepts (eg. Dublin Core), and key metadata notions have been **interoperability, standardisation, discovery, and access** (OLAC, EMELD, Farrar & Langendoen 2003).
- the goals of language documentation mean this is not powerful enough and we need a **theory of metadata**, largely lacking until now



Meta-documentation theory

- **deductive approach:** postulation of axioms and theorems
- **inductive approach:** examination of current and past documentations (so-called ‘legacy materials’) to analyse practices and identify operating principles (as well as lacunae)
- **comparative approach:** examine what other relevant and related fields have done in their meta-documentation, to see what is applicable and what not to documentary linguistics



Deductive: metadata formats so far

- **common or standard:**
 - IMDI (ISLE Metadata Initiative, DoBeS) – rich, for corpus management
 - OLAC (Open Language Archives Community) – compact, for retrieval
 - EAD (Encoded Archival Description), others
- individual organisations (eg. ELAR, AILLA, Paradisec) have developed their own sets and/or allow depositor's own metadata (see Nathan's talk)

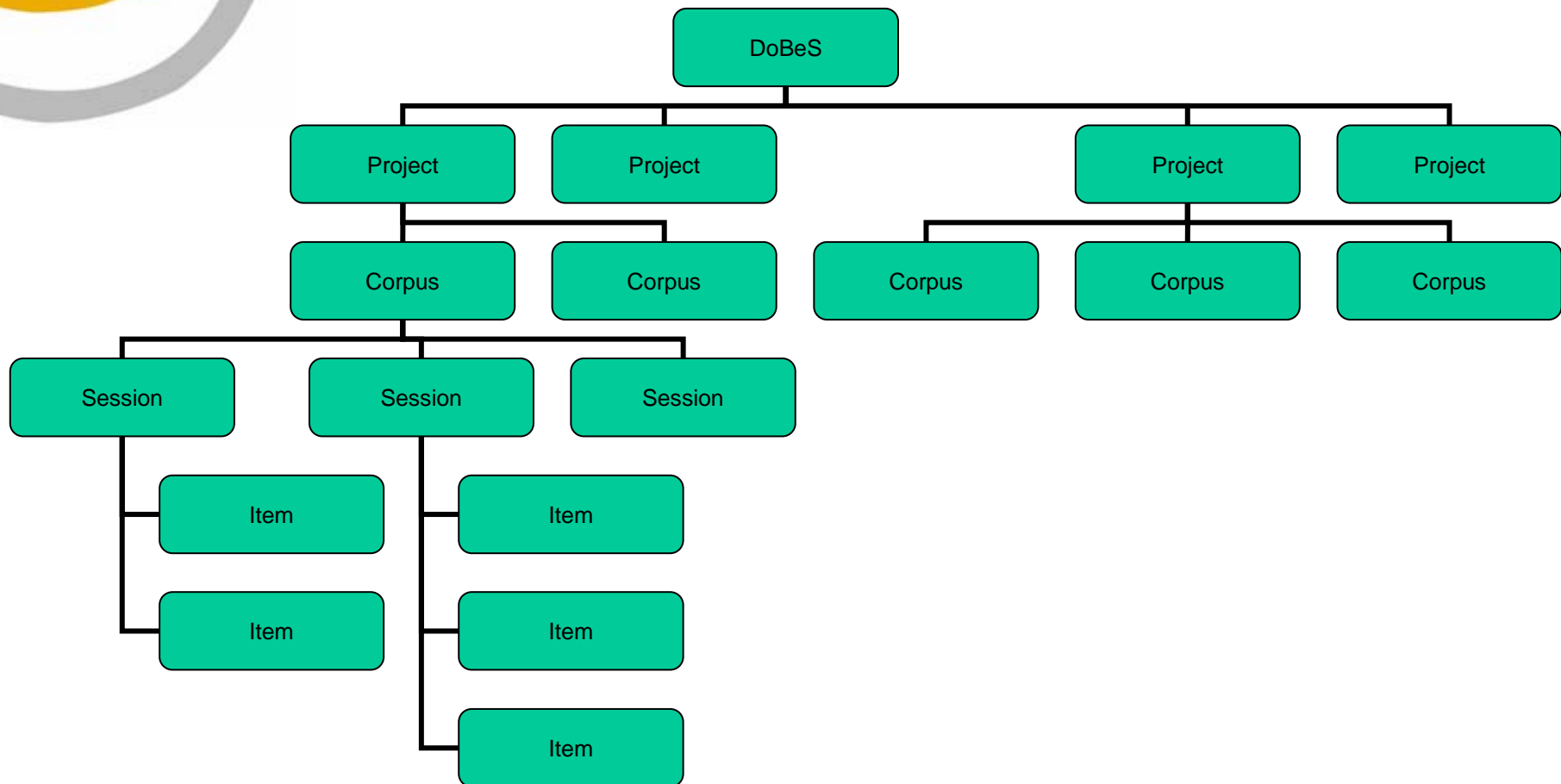


Types of metadata

- creator's / delegate's details
- **descriptive** metadata – content of data
- **administrative** metadata – eg. date of last edit, relation to other data
- **preservation** metadata – character encoding, file format
- **access** and **usage** protocols – eg. URCS
- metadata for **individual** files or **bundles**

- Metadata can apply at various levels

DoBeS model





How do we store metadata?

- **in our heads** – problem: degrades rapidly and not preservable or portable
- **on paper** – problem: not easily searchable or extensible
- **within files** (headers) – problem: not easily searchable or extensible
- **in file/folder names** (eg. SasJBpka09-12_int03.wav – problem: difficult to maintain, breaks easily, not all semantics can be expressed
- **in a metadata system**



Metadata systems

- free text
- structured text (eg. Word tables, XML, Toolbox)
- spreadsheet (eg. Excel)
- database (eg. Filemaker Pro, Access, MySQL)
- metadata manager (IMDI, SayMore)

- or some combination of these that is **usable, flexible** and sufficiently **expressive**



Coming up ...

- **Comparative approach:** metadata in social anthropology (Ember & Good), archaeology (Hanks), and archiving (Conathan)
- **Inductive approach:** metadata practices in archiving (Nathan)
- **Open discussion**
- And don't forget: poster session in Grand Ballroom Foyer on Sunday 9 January from 9:00 AM - 12:00 Noon.