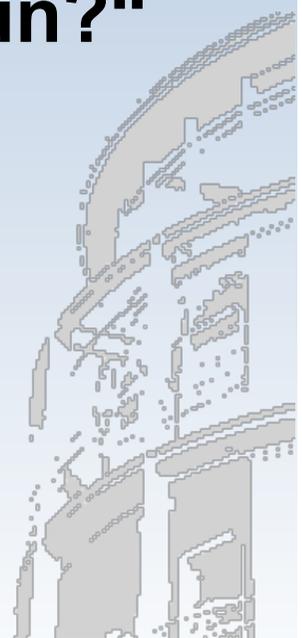




Was können Bibliotheken beitragen? Data Curation Grid

Workshop
„Was haben Bibliotheken mit Grid zu tun?“

Jens Ludwig
ludwig@sub.uni-goettingen.de
Erfurt, 3. Juni 2009





Überblick

- Situation und Rolle der Bibliotheken
- Was ist Data Curation?
- Data Curation Grid und WissGrid



Überblick

- Situation und Rolle der Bibliotheken
- Was ist Data Curation?
- Data Curation Grid und WissGrid



Situationsbeschreibung: Data Driven Science

- Datenmenge nicht (nur) Problem, sondern Chance: Neue Fragestellungen, z.B. nicht nur einzelne Texte, sondern über große Textkorpora.
- Daten sind teuer, aufwändig, z.T. nicht wiederherstellbar: Z.B. Messdaten über Ozonloch einer Polarexpedition.
- Nachnutzung z.T. für unvorhergesehene Zwecke: MashUps von Google Earth und Vogelgrippedaten.



Situationsbeschreibung: Data Driven Science

- Forschungsdaten stehen auf der Agenda aller Forschungsorganisationen in D, EU, USA.
- Bisher waren publizierte Informationen primär Texte. Jetzt zunehmend auch Forschungsdaten und Hybride (DRIVER: “enhance publications”).



Rolle der Bibliotheken

- Neue Publikationsformen kommen auf Bib. zu.
- Datenmanagement wird für Wissenschaft zunehmend wichtiger, da neue Instrumente nicht alle Probleme lösen.
- Bib. existieren schon seit langer Zeit und haben langfristige Sichtweise.
- Bib. “forschen” nicht (eher Meta-Forschung) und sind deshalb im gewissen Sinne Unparteiische.



Was können Bibliotheken beitragen?

- Bib. haben Expertise für Daten-/ Informationsmanagement, z.B. Metadaten
- Bib. können Langfristigkeit, Nachhaltigkeit und Vertrauenswürdigkeit unterstützen/organisieren.
- Bibliotheken haben Gesamtblick: Zwischen verschiedenen Akteuren und Perspektiven im Forschungsprozess vermitteln.
- These: Nicht *für*, sondern *mit* Wissenschaftlern zusammen arbeiten.
- These: “Data Curation” ist eine Hauptaufgabe



Überblick

- Situation und Rolle der Bibliotheken
- Was ist Data Curation?
- Data Curation Grid und WissGrid



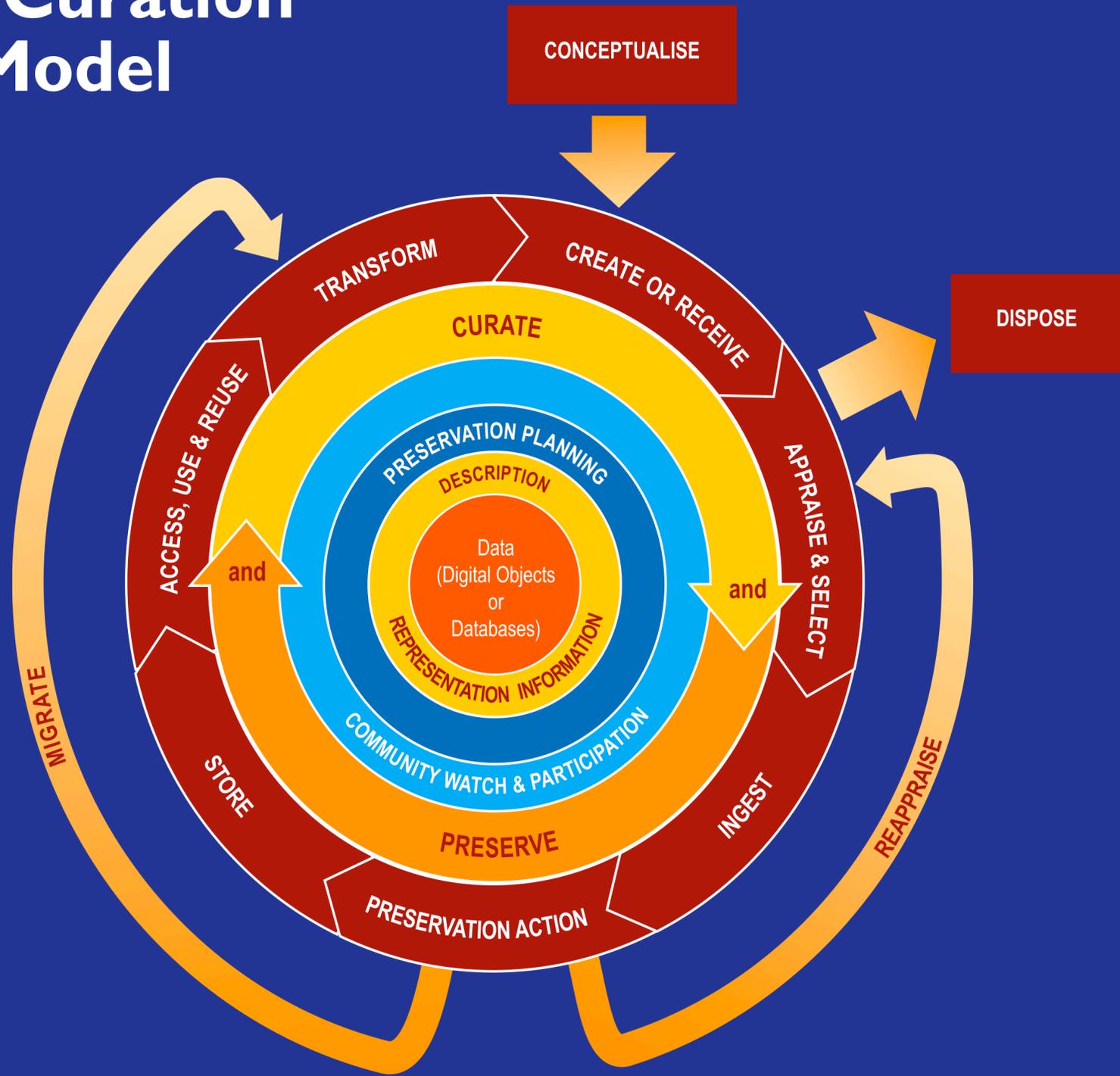
Was ist Data Curation?

"Digital curation [...] is about maintaining and adding value to a trusted body of digital information for both current and future use: [...] it is the active management and appraisal of digital information over its entire life cycle."

Maureen Pennock, Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information,
http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf

Es gibt viele Forschungsprozess und Information-Life-Cycle-Modell, hier: DCC für Curation
<http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf>

The DCC Curation Lifecycle Model

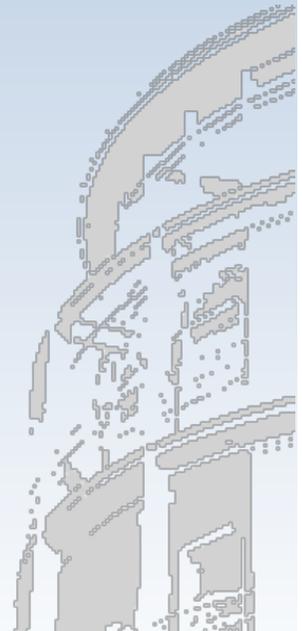




Was sind die Data Curation Aufgaben?

- Description of Data (inhaltlich, technisch, ...)
- Community Watch & Participation
- Preservation Planing

- Conceptualisation of data
- Creation of data
- Appraisal and selection
- Ingest and storage
- Preservation action
- Access, use, and re-use
- Transformations



Der „Unterschied“ zwischen Data Curation und LZA

- LZA wird oft als Wegschließen eines statischen Ergebnisses am Ende missverstanden.
- Gemeinsame Grundfrage: Wie kann ich digitale Objekte in einem anderen Kontext (zeitlich, technisch, fachlich, organisatorisch, ...) *nutzen*?
- „Data Curation“ und „LZA“ sind nur unterschiedliche Betonungen. Die Methoden und Werkzeuge der LZA helfen im gesamten Lebenszyklus.



Überblick

- Situation und Rolle der Bibliotheken
- Was ist Data Curation?
- **Data Curation Grid und WissGrid**



Data Curation Grid: Wieso Data Curation im Grid?

- Im Grid entstehen Daten, werden genutzt und verwaltet, deshalb auch Data Curation sinnvollerweise dort.
- Nachträgliche Maßnahmen sind viel aufwändiger als im Entstehungskontext.
(Nationaal Archief 2005: Nach 10 Jahren kosten Metadaten und Reparaturen das 30-fache)



Data Curation Grid: Wieso Grid für Data Curation?

- Egal ob “Grid” oder “Cloud”: Die Abstraktion von Rechen- und Speicherressourcen ist (meist) ein Fortschritt für Data Curation.
- (Kontext-Unabhängigkeit vs. Kontext-Verlust)
- Instrument für
 - große Datenmengen/Speicherressourcen
 - verteiltes Datenmanagement, wie z.B. für Redundanzen, Bitstream Preservation
 - rechenintensive Aufgaben, z.B. für Konvertierungen, Metadatenextraktion, Validierungen

WissGrid

- Gemeinschaftsprojekt der 5 wiss. Grid-Projekte
- AP 1: Betriebsmodell für akad. Grid-Nutzer
- AP 2: Blaupausen für neue Community-Grids
- AP 3: LZA von Forschungsdaten
 - AP 3.1: Generische LZA-Architektur
 - AP 3.2.1: Modulare LZA-Dienste, zielen auf Interoperabilität. Insbesondere: Identifizierung, Beschreibung, Validierung und Konvertierung von Daten und ihren Formaten
 - AP 3.2.2: Grid Forschungsdatenarchiv



**Vielen Dank.
Fragen?**