The Role of Academic Libraries in the Digital Data Universe

Break-Out Session:
New Partnership Models

Bob Hanisch and Brian Schottlaender
Co-Leaders

Objective

- Develop framework for collective action

- What should we do?

- When?

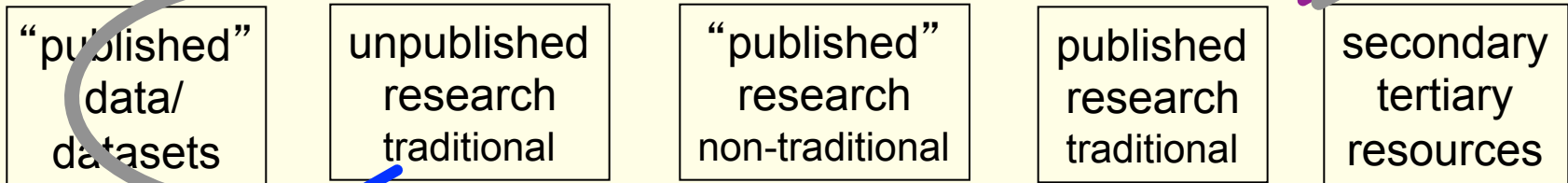- Who should do it?

- Where?

Challenges

- **Overarching**: Crafting partnerships in which the focus is on <u>long-term</u> data stewardship
  - Current partnerships tend to focus on interoperability and integrated access, but lack a long-term component
  - Requires a different kind of institutional commitment and different funding strategies
- **Conundrum**: Short–term, and pressing, need
- Overlapping spheres of responsibility along the research process chain
- Credibility building
- Capacity building
- Changing roles of libraries, and changing <u>perceptions</u> of the roles of libraries (both in- and outside libraries)
  - Current emphasis in libraries is on information discovery, rather than information management (including storage)
  - Libraries need to re-think the partnerships into which they enter (including partnerships with other libraries)
- Definition of the functions comprehended by the word "curation
  - Curation and preservation are not the same thing
  - Preservation is a necessary condition for curation, but not a sufficient one
  - A lot gets preserved (and should) that is not immediately curated
- **Overarching**: Identifying where in the research process chain—or, where in the life-cycle of data—curatorial/preservation activities need to take place
  - Where do partnerships come into play
  - Where are the hand-offs?
  - How do we lower the barriers to participation?

Scholarly Communication

in the past, libraries involved at this end

"traditional" research publication

| "published" data/ datasets | unpublished research traditional | "published" research non-traditional | published research traditional | secondary tertiary resources |

analyzed data/ datasets

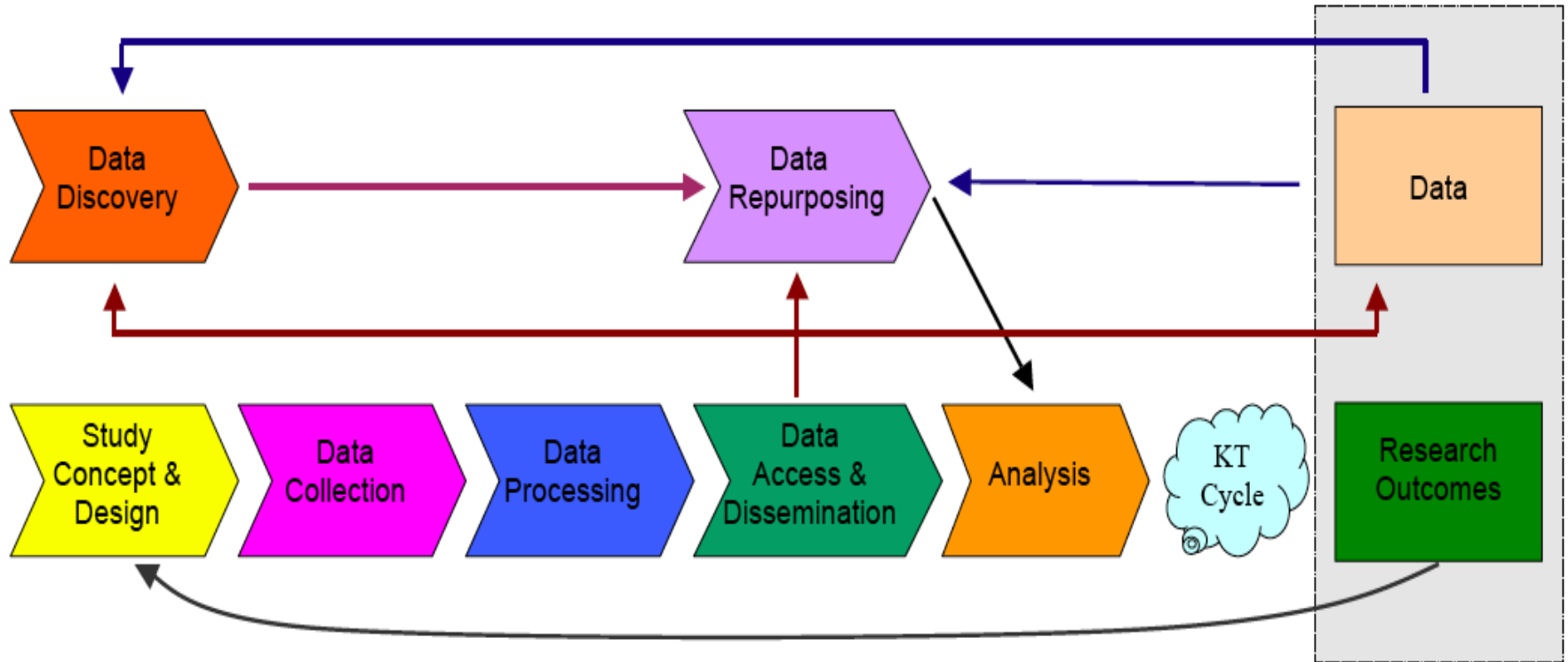currently many attempts to data mine to uncover data…

processed data datasets

metadata curation profiles for data (i.e., data repositories) allow forward/backward movement through scholarly communication process
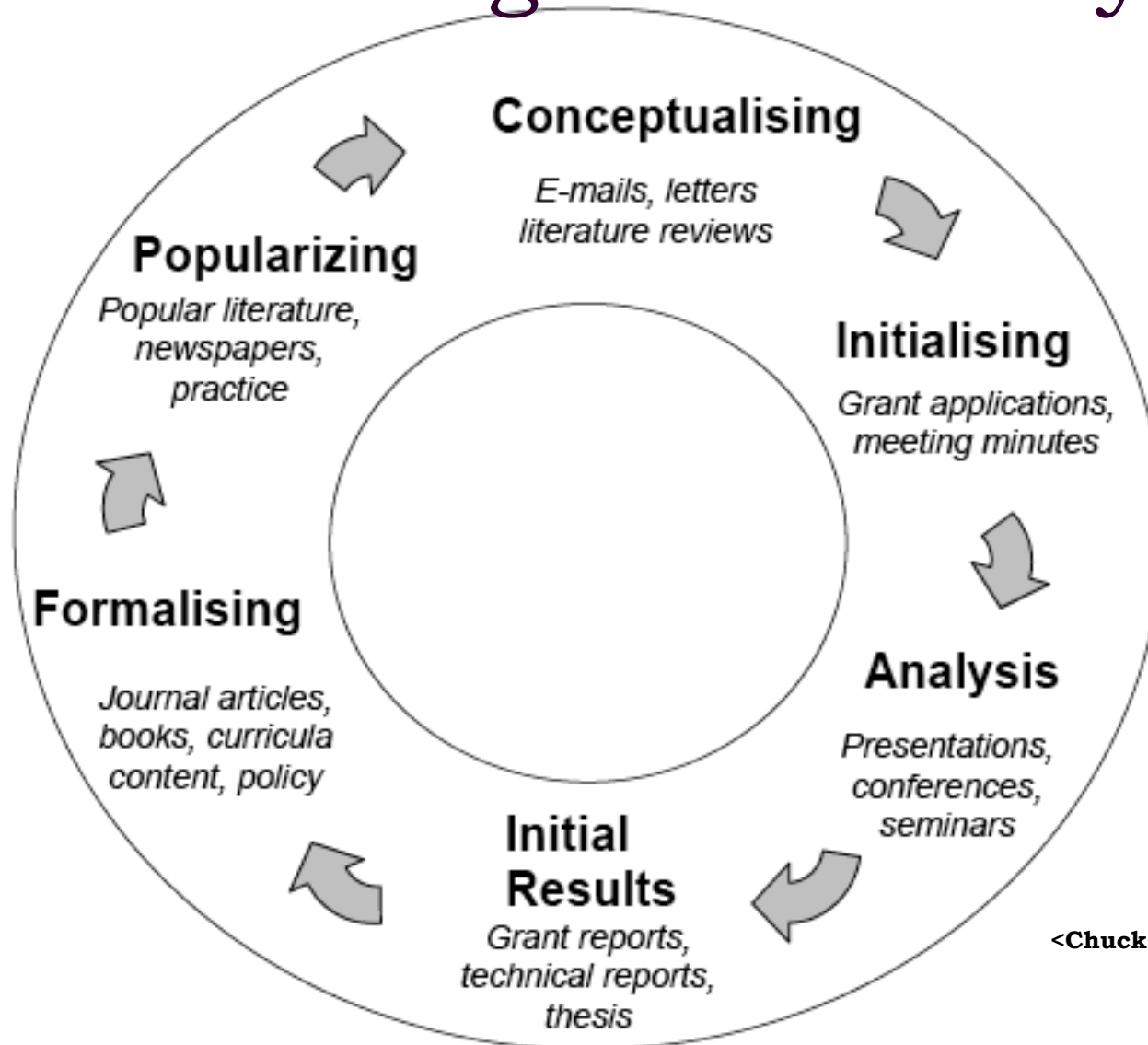
"raw" data/ datasets

<D. Scott Brandt (Purdue)>

The Life Cycle of Research

<Chuck Humphrey (Alberta)>

The Knowledge Transfer Cycle

Conceptualising

E-mails, letters literature reviews

Popularizing

Popular literature, newspapers, practice

Initialising

Grant applications, meeting minutes

Formalising

Journal articles, books, curricula content, policy

Analysis

Presentations, conferences, seminars

Initial Results

Grant reports, technical reports, thesis

<Chuck Humphrey (Alberta)>

Stipulations

- Just as the problem space is a distributed one, so too will the solution space be distributed.
- Long-term stewardship is not about saving bytes; it's about creating, building, and evolving expertise in the community.
- There are multiple players/responsible parties in the problem (and solution) space, who have varying levels of understanding of and interest in the issues:
    - Universities
    - Libraries and librarians (lower-case "l")
    - Domain specialists
    - Computer scientists
    - Standards-setting bodies
    - Editors
    - Professional societies
    - Publishers
    - Commercial and not-for-profit vendors
    - Funding agencies
- "It takes a research <u>community</u> to preserve its data." [emphasis supplied]

Steps to Address Challenges

- Raise awareness and create demand in the research community
- Understand and define the requirements for repositories
  - Granularity
  - Metadata
  - Etc.
- Distribute curation responsibilities across a body of responsible parties roughly equivalent in magnitude (i.e., size, capacity) to the magnitude of the collective data store in need of curation
- Ensure that the work environments of those responsible parties are well supplied with curatorial tools that facilitate their carrying out their responsibilities
- Prototype and test
- Deploy and measure
- Develop business models

Recommendations

1. **Overarching: NSF should facilitate the establishment of a sustainable institutional framework for long-term data stewardship involving the players enumerated above. This framework must:**
   a. Encourage the articulation what, exactly, constitutes "curation" in various disciplines.
   b. Encourage a diversity of designs and approaches that are sympathetic to the needs, practices, and relationships within affected research communities. One size does not fit all.
   c. Encourage the development of distributed partnerships between libraries and research institutions.
2. NSF should fund pilot projects/case studies that demonstrate the intersections between libraries, a limited number of scientific/research domains, and extant technologies bases.
3. NSF should fund projects in which university research libraries develop deep archives of irreplaceable data, assuring descriptions of these data at a minimal level (floor, not ceiling) and facilitating discovery and access to these data, according to prevailing community standards.

   [N.B. in re: 2 and 3 above: It will be important/valuable to find the right balance between prototypes and longer-term commitments.]

Recommendations

4. NSF should require that data management plans submitted as part of the application process identify the players involved in the custodial care of data for the whole of its life cycle, and should support training initiatives to ensure that the research community can fulfill this requirement.

5. NSF should foster the training and development of a new workforce in data science
   a. Promote new curricula
   b. Develop new programs
   c. Link to training of domain scientists and information/ library scientists

6. NSF should partner with IMLS to train information and library professionals (extant and future) to work more credibly and knowledgably on data curation as members of research teams