Conceptualizing Worksets for Non-Consumptive Research

Jacob Jett, Chris Maden, Colleen Fallaw, Megan Senseney, J. Stephen Downie
Center for Informatics Research in Science & Scholarship, Graduate School of Library & Information Science, University of Illinois at Urbana-Champaign

References:

WCSA is funded by The Andrew W. Mellon Foundation, Grant Ref # 21100666
Special thanks to Timothy W. Cole and Allen H. Renear.

Background
HathiTrust Digital Library (HTDL):
- Comprises over 13.2 million volumes
- Containing 4.6 billion pages of digitized content
HathiTrust Research Center (HTRC):
- Collaborative effort between the University of Illinois and Indiana University
- Developing a rich digital environment that facilitates scholarly use of the HTDL's corpus within a non-consumptive research paradigm

The Non-Consumptive Research Paradigm
- Two thirds of HTDL corpus is still under copyright
- The HTRC’s Workset concept helps by providing:
  - A selection & aggregation tool for researchers
  - An access & interaction tool for algorithms

Functional Requirements for the HTRC Workset Model

A workset in the HTDL context:
- is a container for a scholar’s aggregated units of analysis - analogous to a scholar’s research collection;
- is a persistent globally unique entity that can be directly cited;
- possesses provenance properties supporting change awareness within the HTRC’s architecture so that a description of its nature at the time of analysis persists over time;
- is flexible enough to allow for the aggregation of heterogeneous resources, with regard to:
  - granularity and
  - source repository; and
- derives the values for a number of its properties (i.e., its metadata) from the values of certain properties of its constituent members that propagate allowing it to support of various types of filtration (cf. Wickett et al., 2010; Wickett, 2012).

Bibliographic resources in the HTRC context:
- are persistent globally unique entities that can be directly cited, and
- possess provenance properties supporting change awareness.

Future Work: Some Low-hanging Fruit

Improving Provenance: Disambiguating Copies

Future Work: More Advanced Concepts

Selecting Specific Chapters

Bibliographic Granules: Page-level Entities

Future Work: Some Low-hanging Fruit

De ludo scacchorum
- published by: De ludo scacchorum
- publishedAt: 1474
- placeOfPublication: Bruges
- language: Latin

William Caxton
- born: 1288
- died: 1401
- placeOfDeath: London
- language: English

Jacques de Cessolis
- born: 1288
- died: 1322
- placeOfDeath: Bruges
- language: French

University of California
- hasRepresentations
- hasParts
- hasDescription

[The Game of Chess]

EER0-TCP
- publishedAt: 1474
- publisher: EER0-TCP
- language: English

Benjamin Franklin
- born: 1706
- died: 1790
- placeOfBirth: Boston
- language: English

ES ödeme
- language: Polish

Bibliographic Resources

Metadata (in Named Graph Form)

Worksets

Metadata (in Literal Form)

Color Key: