Engaging with RDA

To improve data management and sharing practices in the Humanities

Bridget Almas
RDA Webinar
27 November 2017
Personal Background and Motivation

- Background in commercial software development

- Became engaged with RDA in 2013
  - While working for the Perseus Digital Library, and subsequently the Perseids Project, at Tufts University
  - Maintaining and developing infrastructure for digital humanities research, pedagogy and data sharing

- Motivated by the assumption that infrastructure for the sciences could serve the needs of the humanities

@BridgetAlmas
From Reinvention and Redundancy...

... to Shared Understanding and Collective Improvements

Image source: https://twitter.com/davidjbland/status/634059970699358208

@BridgetAlmas
Common Needs - Services & Standards

- Persistent Identifiers
- Collection Management
- Repositories
- Data Registries and Formalized Data Types
- Metadata Management
- Content Management
- Provenance and Reproducibility
- Search and Indexing
- Peer Review/Workflow
- Brokers

@BridgetAlmas
Common Needs - Workflow Tools

- Curation
- Annotation
- Analysis
- Visualization
- Publication
Common Challenges

- Governance
- Sustainability
- Project-based funding models
- Extensibility
RDA Participation ...

… Getting Out What You Put In

@BridgetAlmas
Roles held within RDA

- Technical Advisory Board 2013-2015
- Co-Chair of Research Data Provenance IG 2013-2016
- Co-Chair of Data Fabric IG 2016-2017
- Co-Chair of Research Data Collections Working Group 2016-2017
- Member of RDA-US Leadership Committee 2014-2017
Gained from Participating in RDA

- Cross domain perspective
- Solutions
- Challenges
- Colleagues and collaborators

@BridgetAlmas
Ways to participate

- Adopt an output
- Come to the plenaries
- Follow discussions of the Interest and Working Groups
- Join an existing Interest Group
- Join an existing Working Group as an adopter or use case provider
- Start a Working Group
  - Consider doing your project based work *within the context of a Working Group*
Putting the pieces together
End users, developers, and automated processes deal with persistently identified, virtually aggregated digital objects, including collections which are overlays on multiple network services which in turn are overlays on existing or future information storage systems.

Source: Larry Lannom, CNRI, Data Fabric IG
Perseids Use Case: Workflow

Stage 1: User creates data. It goes in the User Collection and the Subject Collection.

Stage 2: User adds data to a community. It gets added to a Community Collection.

Stage 3: User requests peer review. It gets added to a Review Board Collection.

Stage 4: Data passes peer review. It leaves the review board collection and gets added to a Published Collection specific to the data type.

http://perseids.org

@BridgetAlmas
Perseids Use Case - Why

Enabling preservation and retrieval of all data...

- created by User X
- approved by Community X
- of a specific type about a specific canonical resource
  - All treebank data for Homer’s Iliad Book 1 Lines 1-10
  - All translation alignments of Homer’s Iliad Book 1, Lines 1-10
  - All semantic annotations on Vergil’s Aeneid Book 1
  - And so forth….

@BridgetAlmas
Perseids Use Case - What

Data objects which can be:

- Persistently identified
- Consistently typed
- Machine Actionable
- Reproduced
- Attributed
- Cited

@BridgetAlmas
Perseids Use Case - How

User
- create data obj

Perseids
- query(datatype properties)
  - datatype ID
- create item(datatype id)
  - collections?filter=(datatype=datatypeid, access=writebyuser)
- (collection list)
  - collection

PIT API
- collections

DTR API
- collection(id)/matches(item)
  - (items)
  - modeltypes(collection, modeltype)

Collections API
- model allows duplicates?
- add or edit?
  - add
- pid

POST collections/(id)/items(item)
  - create(item)
  - pid

@BridgetAlmas
Perseids Use Case - Vision

@BridgetAlmas
Alpheios Project: use cases for RDA Outputs

- Crosses domains: pedagogy, linguistics, neuroscience, classics, computer science
- Non-profit, fully open source, open data
- Will be collecting data we want to share, reusing data from other sources and domains
- Expect to make use of (at least):
  - Collections API
  - Data Types Registry
- But also, the RDA community

http://alpheios.net @alpheios

@BridgetAlmas