

DSpace@Cambridge

Planned Digital Preservation R&D



BL/CURL/DPC Forum
19 October 2004



This Presentation

- Context
 - [DSpace@Cambridge](#) Repository policies and trends
- Digital Preservation R&D
 - Representation Information
 - Preservation Planning
 - Provenance / Change Information
 - File Format Investigation (ArcInfo Export format)
- Other R&D



BL/CURL/DPC Forum
19 October 2004



Context :: Repository Policies

- No mandates on
 - Type of material
 - File formats
 - BUT actively advice on good practise
- Contributors
 - All voluntary
 - Some we found
 - Some found us



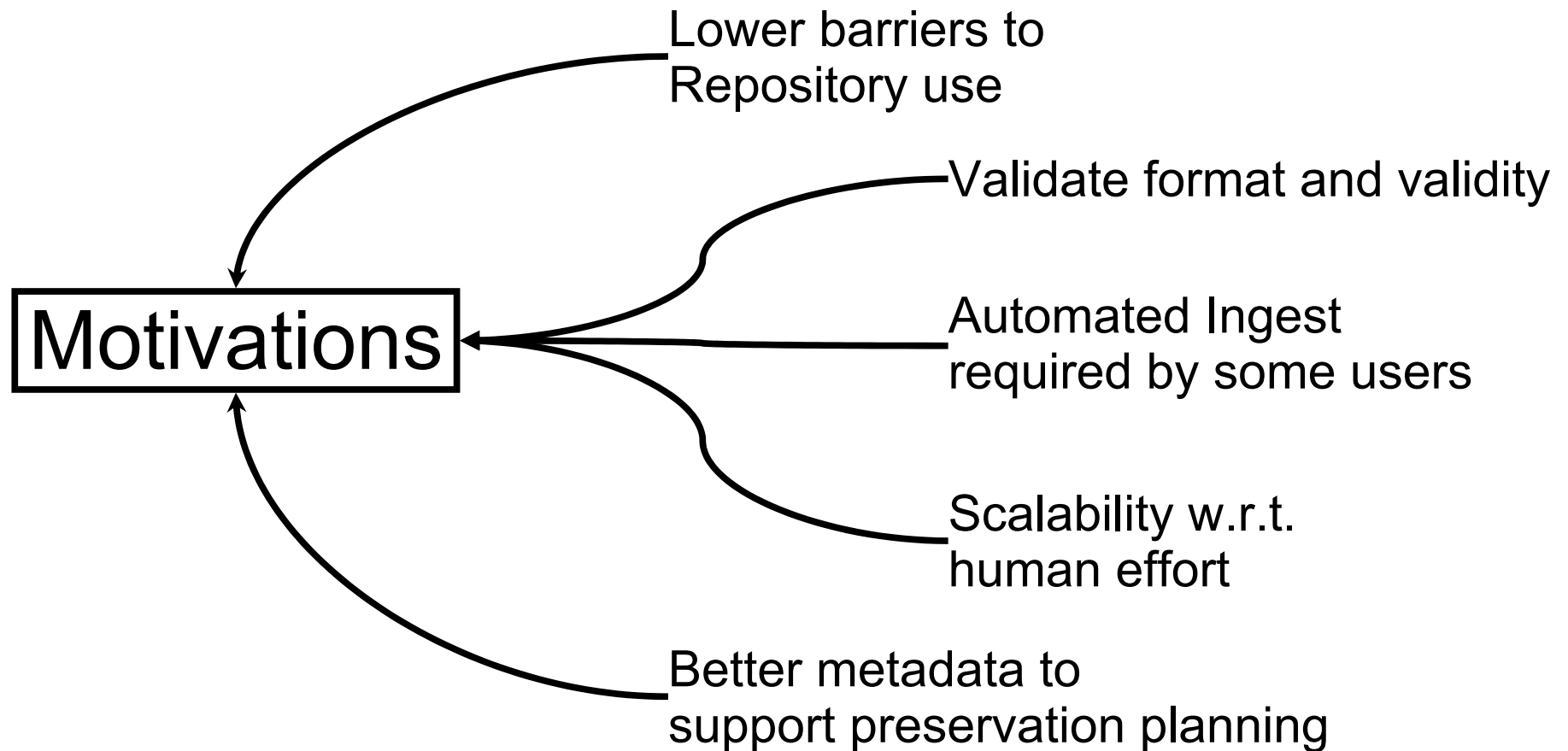
BL/CURL/DPC Forum
19 October 2004



Context :: Trends

- Comparatively few research papers
 - 2 preprints and 164 working papers in 616 items (180 GB) (2004-10-01)
- Lots of digital video and images
 - Users would understandably rather use our DSpace service than build and admin their own storage
- Encouraging variety of formats and domains
 - From molecules to Monet
 - From XML to digital video

Digital Preservation :: Representation Information



DSpace Type Registry Integration

- Finer grained than MIME type
- Format + version (+ profiles?)
- Minimal Interpretation of “Type Registry”
 - Deal with representation issues separately
- Alternatives for initial implementation: -
 - GDFR (if announced soon)
 - PRONOM
 - TOM

Automated Processes

- Object Identification
 - Integrate JHOVE
- Migration
 - TOM
- Metadata Extraction
 - JHOVE
 - RLG Automatic Exposure
 - GNU tools proof of concept (e.g. file, imagemagick)

Preservation Planning

- Encoding of human-understandable basic action plans
- Enable sharing of action plans
- Evolution of plans e.g. to add automation
 - Retain human validation of automated steps
- Plans share prototypical bitstream roles
 - e.g. notion of “current normalized version” can form part of several action plans
 - Separation of concerns

ArcInfo Export Format Investigation

- Using GIS material from Cambridge Dept of Archaeology and MIT
- Migration Issues
 - Assess available Open Source tools
 - Vendor support from ESRI
- Emulation options
- Current best practice
- Vendor / product information

DSpace Developments

- DSpace Developments
 - Currently records all item / metadata changes
 - Difficult to scale with # items
 - Refine policy and implement
- URL handling – dual approach
 - Semantically rich URLs for current use
 - Semantically void persistent identifier actioning URLs for preservation purposes

Potentially Useful Information

Jim Downing - ojd20@cam.ac.uk

Dspace@Cambridge - <http://www.dspace.cam.ac.uk>

Some DP related discussion on - <http://wiki.dspace.org>

DSpace Digital Preservation mailing list -
<http://mailman.mit.edu/mailman/listinfo/dspace-preservation>



BL/CURL/DPC Forum
19 October 2004

