

DSPACE

DSpace & Preservation

Chris Yates

Aberystwyth University

RSP Technical Support Officer

- RSP Technical Repository Support Officer
- Based at Aberystwyth University
- Background in VLEs and Institutional Repositories
- DSpace Developer & Trainer



- An open source **digital asset management system**
- A technology platform for **Institutional Repositories**
- A **federation** of digital repositories across multiple academic research institutions
- A **production service** of the Institutional Libraries to its local research community

- Repositories don't “do” preservation alone
- Preservation operations are defined by
 - Digital collections in hand
 - Content? (Software, Text, Images, Audio, Video)
 - Cost/benefit tradeoffs
 - Organizational Budget
 - Local policy
 - Current information strategy – what has the Organisation
- Repositories “can” enable preservation policy and practice

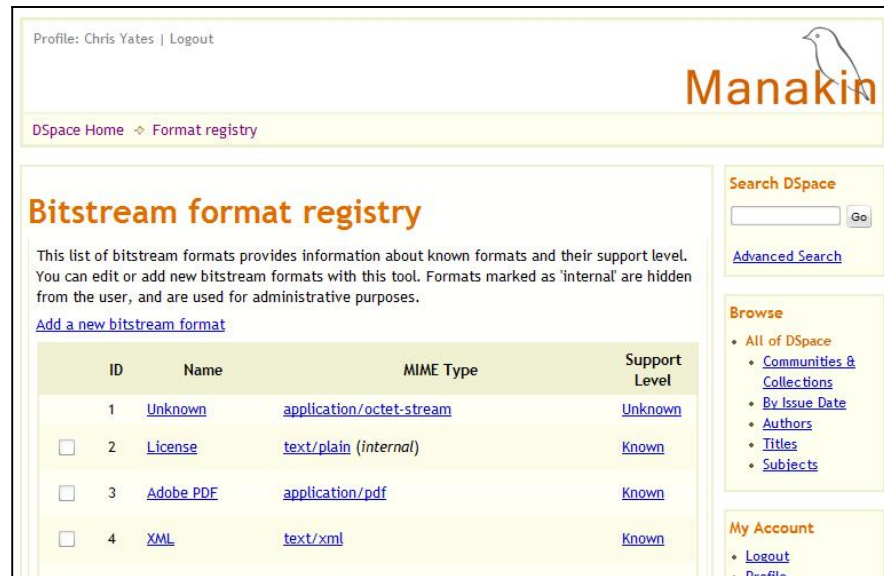
- DSpace features to aid preservation
 - Format Registry
 - Bit Integrity Checker
 - Item History (Audit Trail)
 - Logs
 - Handles
 - DSpace Foundation
 - The Future



(1)

- Preservation of digital objects has specific demands of formats
 - Validation
 - Obsolete Format Detection
 - Migration

- Specification of supported formats in IR

A screenshot of the 'Bitstream format registry' web interface. The page has a header with 'Profile: Chris Yates | Logout' and a 'Manakin' logo. Below the header is a navigation bar with 'DSpace Home' and 'Format registry'. The main content area is titled 'Bitstream format registry' and includes a description: 'This list of bitstream formats provides information about known formats and their support level. You can edit or add new bitstream formats with this tool. Formats marked as 'internal' are hidden from the user, and are used for administrative purposes.' There is a link 'Add a new bitstream format'. A table lists four formats: 1. 'Unknown' with MIME Type 'application/octet-stream' and Support Level 'Unknown'; 2. 'License' with MIME Type 'text/plain (internal)' and Support Level 'Known'; 3. 'Adobe PDF' with MIME Type 'application/pdf' and Support Level 'Known'; 4. 'XML' with MIME Type 'text/xml' and Support Level 'Known'. On the right side, there is a 'Search DSpace' section with a search box and a 'Go' button, and a 'Browse' section with links to 'All of DSpace', 'Communities & Collections', 'By Issue Date', 'Authors', 'Titles', and 'Subjects'. At the bottom right, there is a 'My Account' section with links to 'Logout' and 'Profile'.

- Identification of unsupported formats in the IR
- Formats can also be validated through external services (e.g. PRENOM / JHOVE)

- Bit Rot - The decay of physical storage media

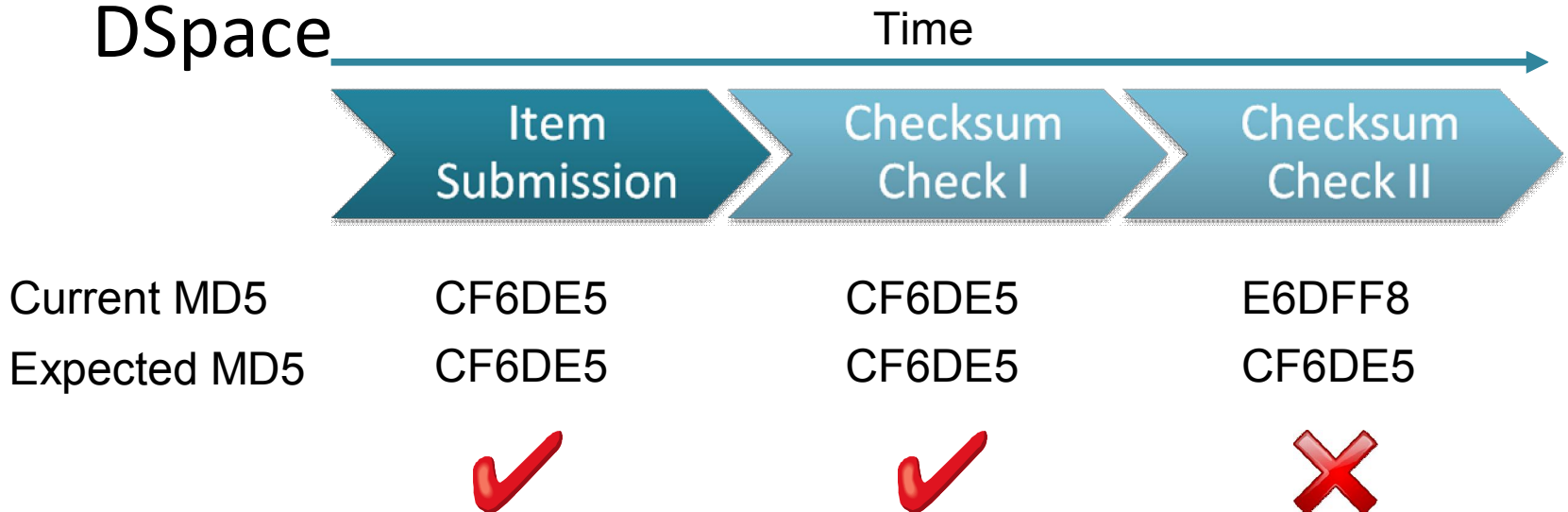
10111**1**01 $\xrightarrow{\text{Time}}$ 10111**0**01

- State of original file can change with time
 - Manual editing
 - Corruption
 - Virus



- Solution -> The utilization of a **checksum**

- Checksum -> fixed datum computed from a block of digital data
- Checksum script can be scheduled to verify and report the integrity of each item in DSpace

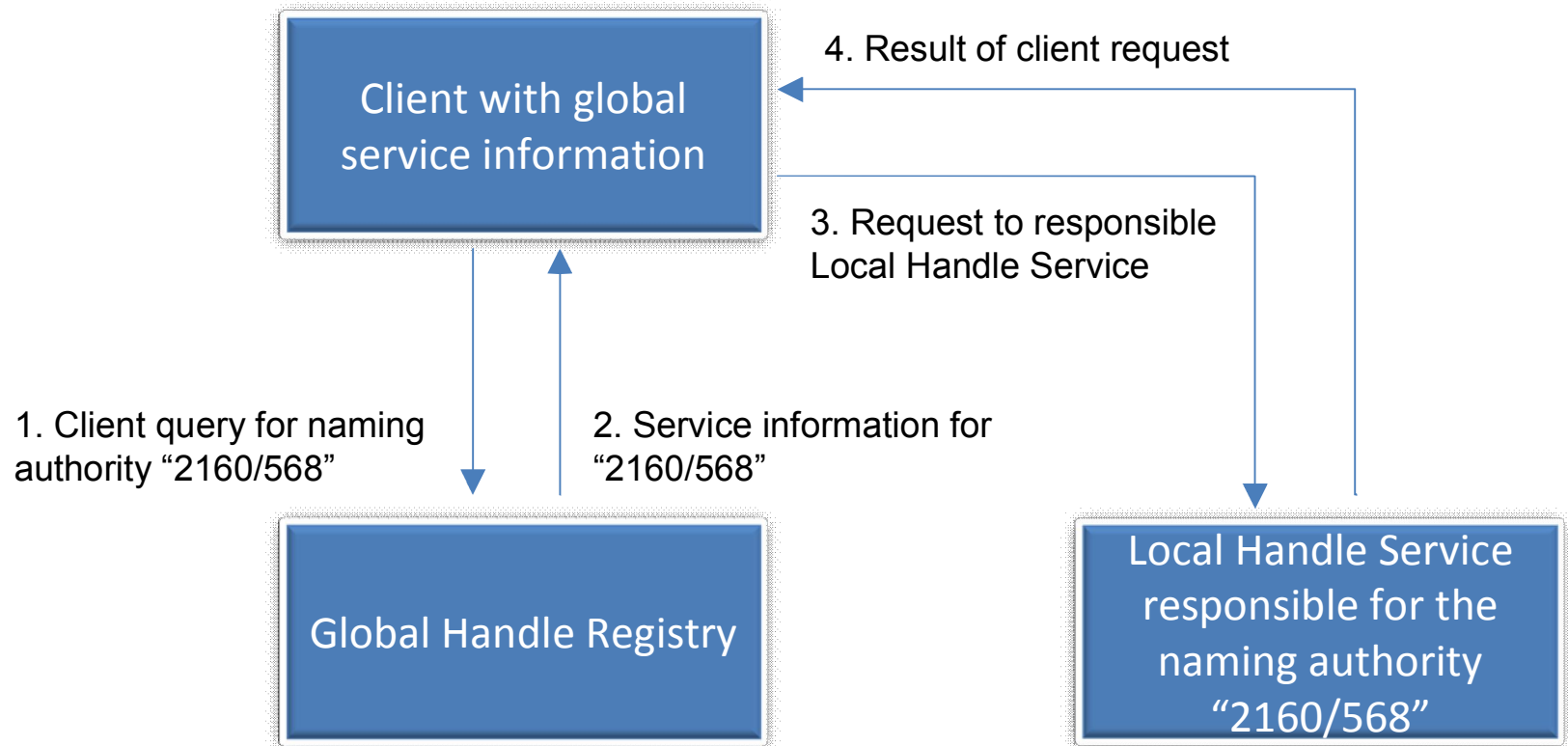


- Handles allow the resource to be uniquely identified in a way that will not change if the resource is renamed or relocated
- Resource address is identified by a unique handle assigned by a common registration service

<http://hdl.handle.net/2160/568>

Registration Service	Handle Prefix	Local Identifier
http://hdl.handle.net	2160	568

- Handle resolution: <http://hdl.handle.net/2160/568>



- Logs provide a record of events within DSpace

A screenshot of the DSpace 'Edit Item' interface. It shows tabs for 'Item Status', 'Item Bitstreams', 'Item Metadata', and 'View Item'. The 'Item Metadata' tab is active, displaying a form to 'Add new metadata'. The form has fields for 'Name' (a dropdown menu showing 'dc.contributor.advisor') and 'Value' (a text input field). There is also a 'Language' dropdown and an 'Add new metadata' button. To the right, there is a 'Search DSpace' section with a search bar and a 'Go' button, and a 'Browse' section with links to 'All of DSpace', 'Communities & Collections', 'By Issue Date', 'Authors', 'Titles', and 'Subjects'.

- Log Extract:

2008-12-07 18:06:42,224 INFO

org.dspace.content.Item@csy@aber.ac.uk:session_id=D2F63D5B508FE5D1B:ip_addr=127.0.0.1:update_item:item_id=1236



- Provides provenance information for items
- Captures a time-based record of significant changes in DSpace
- RDF Data model suitable for later 'refactoring'



More Information:

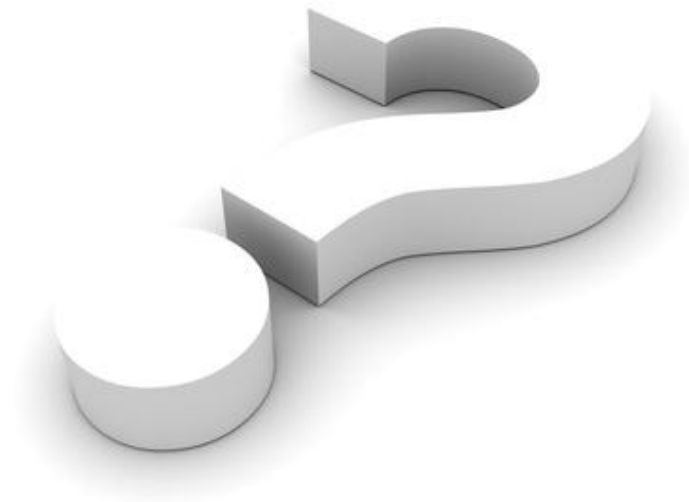
<http://wiki.dspace.org/index.php/HistorySystemPrototype>

- Why the DSpace foundation?
 - Providing leadership and support
 - Drive & co-ordinate development
 - Provide stability and longevity
 - Promote wider distribution and use
- Ensuring the future and continued development of the DSpace software

- DSpace 2.0
 - JHOVE2 (Format Recognition)
 - Remodelling of storage architecture
 - DuraSpace (Fedora/DSpace collaboration)
Distributed Storage Layer
 - External Identifiers
 - Flexible Metadata



Thank you for listening!



- (1) Female hands and tree
(<http://www.fotolia.com/id/10616058>)
HP_Photo - Fotolia.com
- (2) Flash drive in hand isolated on white background
(<http://www.fotolia.com/id/10315587>)
HP_Photo - Fotolia.com
- (3) Female hands holding a small plant
(<http://www.fotolia.com/id/10616376>)
HP_Photo- Fotolia.com
- (4) White Question Symbol
(<http://www.fotolia.com/id/8017330>)
Eugene Kuklev Fotolia.com
- Business Models related to Digital Preservation, ERPANET Seminar. Julie Walker MIT September 20-22, 2004