



Data Lifecycle & Preservation Process

PIN Workshop
Paris , 4th Jan 2013

Danièle Boucon (CNES) and Esther Conway (STFC)



Digital Preservation Landscape



What we need to Preserve

The LTDP Preserved Data Set Document (PDSC) mandates the type of information to be preserved for the following stages

- **Pre Mission**
- **Mission Implementation**
- **Mission Operations**
- **Post Mission**

ESA UNCLASSIFIED – For Official Use



DOCUMENT

Long Term Data Preservation

Preserved Data Set Composition

LTDP/PDSC

LTDP Guidelines (How we intend to do this)

Theme 1: Preserved Data Set Composition

Theme 2: Archive Operation and Organization

Theme 3: Archives Security

Theme 4: Data Ingestion

Theme 5: Archive Maintenance

Theme 6: Data Access and Interoperability

Theme 7: Data Exploitation and Re-processing

Theme 8: Data Appraisal and Purge Prevention



LONG TERM PRESERVATION OF EARTH
OBSERVATION SPACE DATA

EUROPEAN LTDP COMMON GUIDELINES

LTDP Guidelines (Preservation analysis workflow for EO space data)

Definition of Preservation Objective and Designated Community

Definition of Preserved Data Set Content

Creation of PDSC Inventory

Risk Assessment

Preservation Planning and Implementation

Risk Monitoring and Asset Evolution

Risk Analysis

- **Technical Risk**
- **Semantic (Knowledge base)**
- **Acquisitional**
- **Organisational**
- **Resourcing and Skill base**
- **Identification/Validation**
- **Community**

Types of Preservation Action

- **Risk Acceptance and Monitoring**
- **Citation**
- **Description and Information Capture**
- **Software Capture & Emulation**
- **Transformation**

Quality Assurance and Testing

Trust occurs when the archive appraises a solution as satisfactory for one of the following reasons

- **Trust in a custodial organization;** when the archive relies on an external organization to maintain the integrity and supply of important information. The accepted reputation of the organization supplies the required assurance.
- **Trust in a standardization process;** when an archive acquires descriptive information which has produced as a result of a standardization process such as ISO.
- **Quality of Sources;** this occurs when an external organization supplies an archive with an information object. Trust is based upon the belief that the supplier has delivered a quality preservation solution.

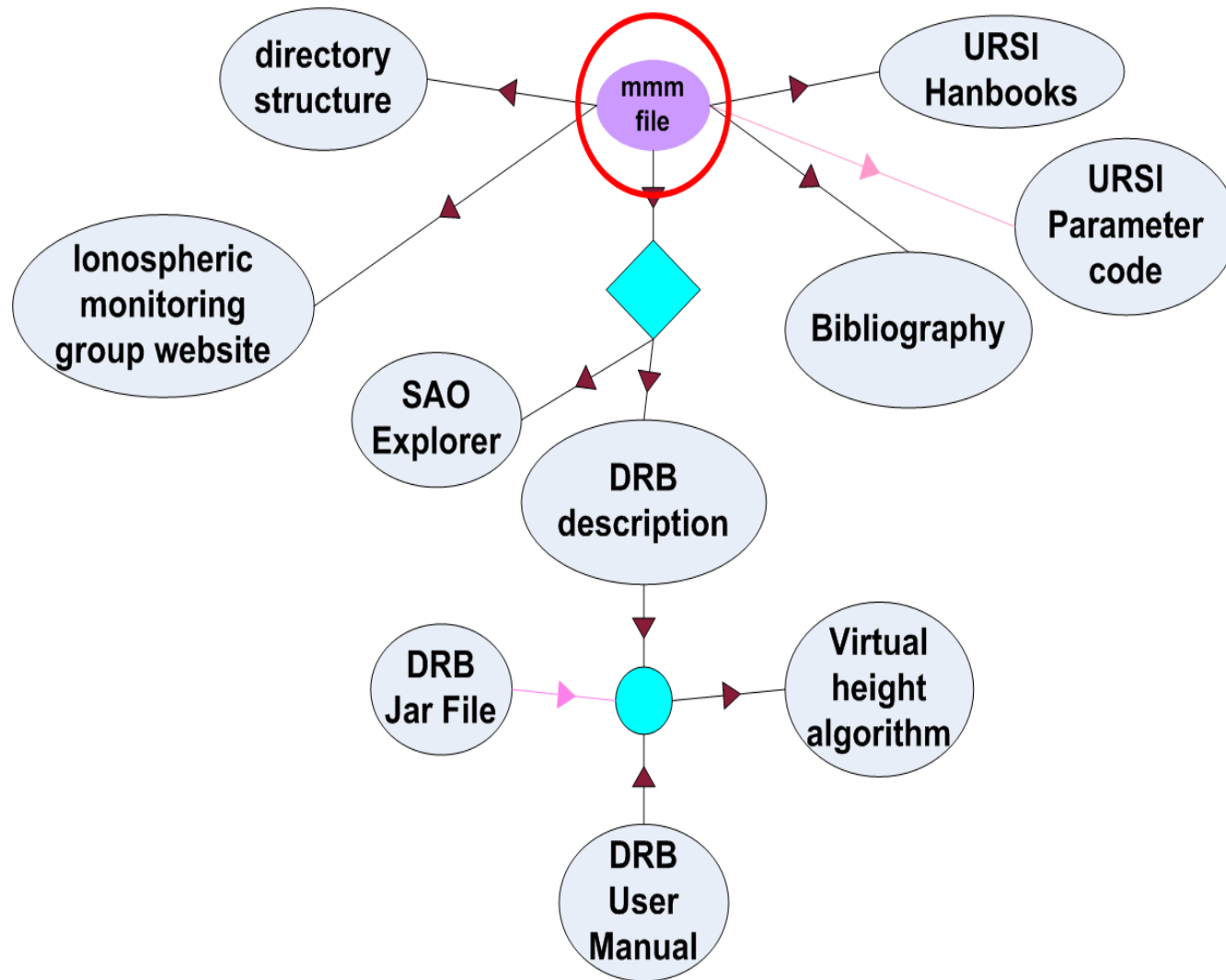
Testing

- **When an archive cannot fully trust a solution it must then employ testing to gain necessary assurance. We consider three testing scenarios.**
- **Passive testing;** occurs when a preservation solution is exposed to an active user community with the expectation that they will report any deficiencies.
- **Proactive testing;** occurs when external experts are invited to test a preservation solution.
- **Direct testing;** occurs when the archive conducts testing itself.

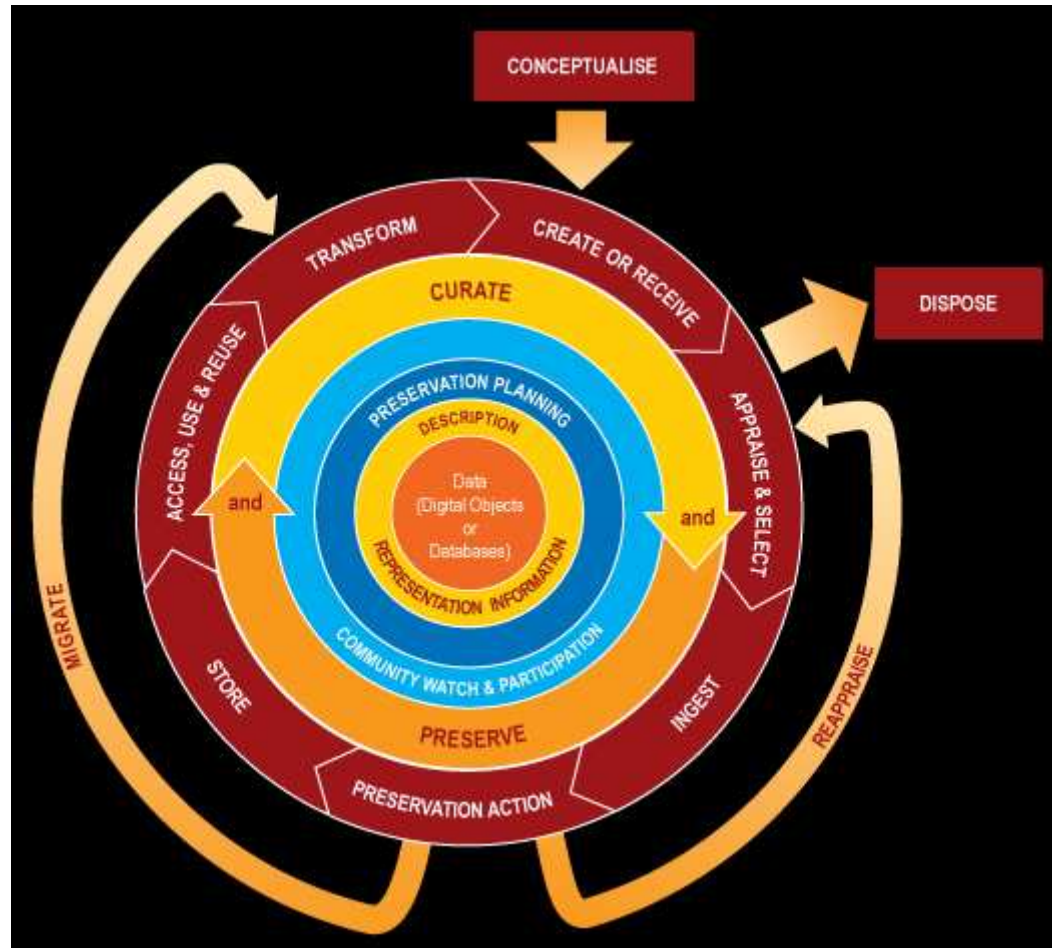
LTDP Evolution of the preservation process

No preservation solution is permanent. Long term preservation requires the evolution of these types of models in order to accommodate changes in the following

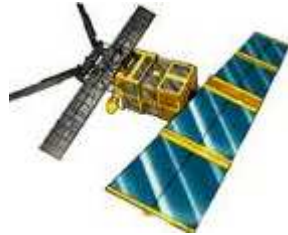
- **Preservation Objective**
- **Designated Community**
- **Realisation of Risks**



DCC Lifecycle Model



Related LTDP Work 2013



- **Produce a Risk Analysis for EO data sets**
- **Conceptual Model for the “Preservation State” of archived EO data sets**
- **Application Case to ERS1 ATSR satellite data and a climate data test case**
- **EO LTDP Preservation Process Model and formal definition of key preservation states**