Preserving 3D data:

Best Practices from a UK perspective

Kieron Niven
Digital Archivist, Archaeology Data Service

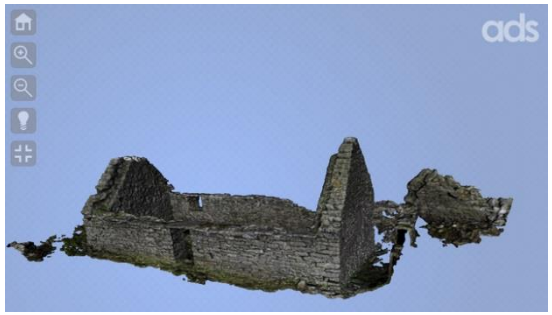
5th February 2018

Overview

1. An Overview of the ADS and the UK heritage and preservation landscape

2. Guides to Good Practice and other guidance

3. 3D data: what is it and how do we preserve it?



An overview of the ADS

The Archaeology Data Service

- Set up in 1996 at the University of York
- Core aim: the long-term digital preservation and dissemination of data
- **Research data** archives
- **Development-led** (commercial) fieldwork archives
- Primarily UK focussed
- Advice to Research Councils and National Heritage Agencies

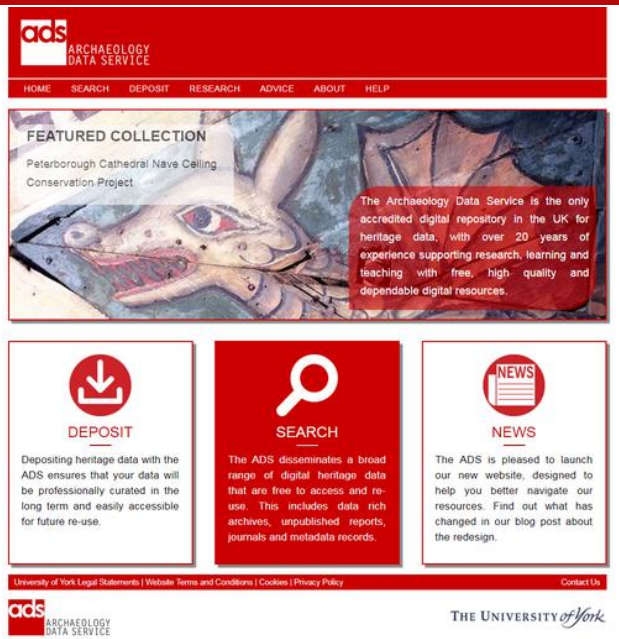
Other UK bodies and groups

- National bodies: Historic England, Historic Environment Scotland, RCAHMW.
- Bedern Group
- Forum on Information Standards in Heritage (FISH)
- Digital Preservation Coalition

Cooperation and involvement within and beyond your sector is key.

What do we archive?

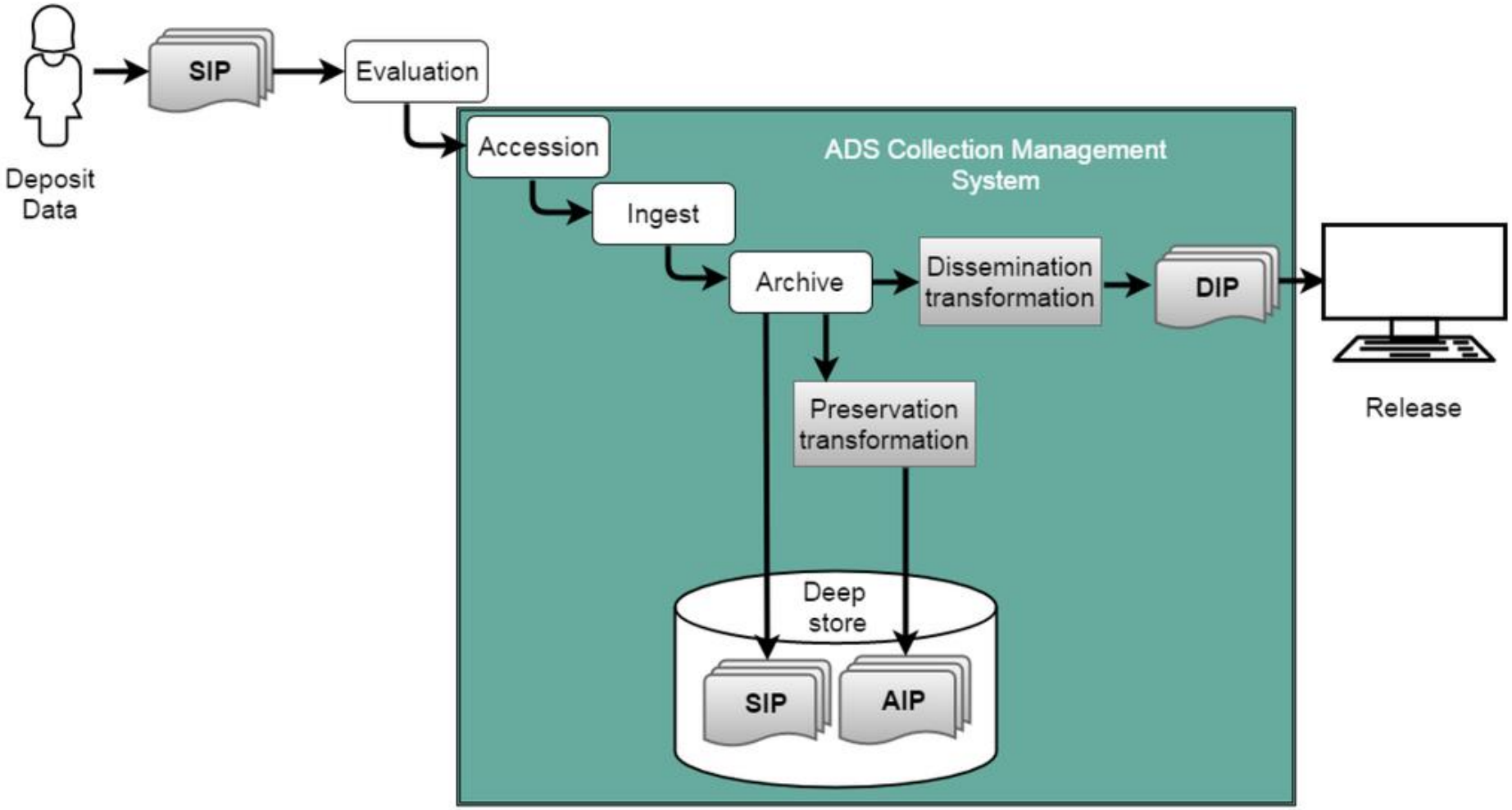
ADS collections:



- around 2.5 million files
- around 17TB of data
- c.1000 ‘rich archives’
- largely image data (TIFF, JPG)
- …and then PDF files (e.g. reports, journals) and DOC files
- databases, spreadsheets, geophysical survey data, GIS, CAD, video

ads ARCHAEOLOGY DATA SERVICE

How do we archive it?



Overall Aim

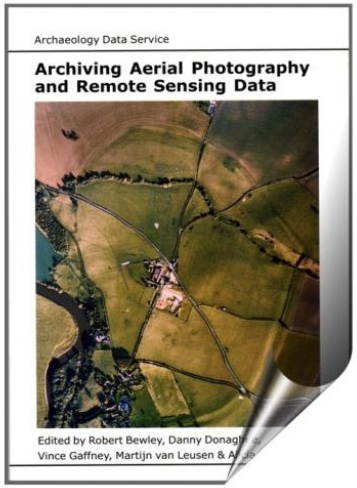
- Ensure that digital data is successfully archived, managed, and accessible in a digital format
- Preserve data, through normalisation and migration, in standardised formats to ensure long-term accessibility
- Ensure data is properly documented and understandable
- Documented in our Preservation Policy and Repository Operations documents:
  - http://archaeologydataservice.ac.uk/advice/PreservationPolicyRev.xhtml
  - http://archaeologydataservice.ac.uk/advice/RepositoryOperations.xhtml

Guides

Guides to Good Practice:
Initial stage of development (hard copy guides):

1998 GIS: A Guide to Good Practice
1998 Archiving Aerial Photography and Remote Sensing Data
1999 Digital Archives from Excavation and Fieldwork (+ Rev. 2nd Ed.)
2001 Geophysical Survey Data in Archaeology
2002 CAD
2002 Creating and Using Virtual Reality:
a Guide for the Arts and Humanities

Hard copy and online, covered similar core elements.



Role of the Guides

- **File formats**
  - how they are used
  - which are best suited to long-term preservation and access
  - preference for non-proprietary formats and open standards, uncompressed formats, formats which use plain text and are human readable
- **Metadata and documentation**

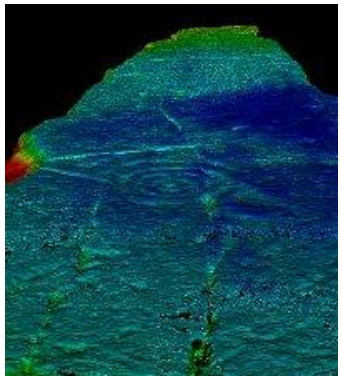
A Brief History of the G2GPs

Phase of Project-based development:

2005-6 The Big Data project:

- Wessex Archaeology *Wrecks on the Seabed* (mag., sub-bottom, sidescan, multibeam)
- Durham University *Breaking Through Rock Art* (laser scan)
- English Heritage *Where Rivers Meet* (lidar)
- Project produced a final report and a set of recommendations for future research – all available online:

http://archaeologydataservice.ac.uk/research/bigData.xhtml

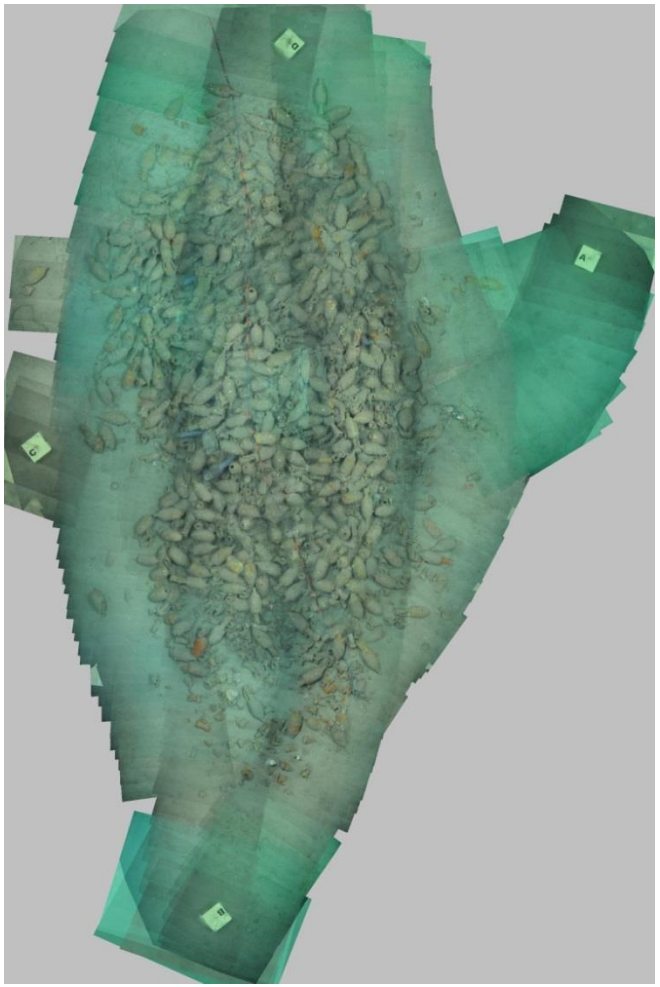


A Brief History of the G2GPs

More project-based development…:

2006-9 Virtual ExploratioN of Underwater Sites (VENUS) (with INRAP):

- Aimed to develop methodologies and technological tools for the virtual exploration of underwater sites (using UAVs).
- ADS deliverables (again, online):
  - Exemplar archive (images, multibeam & sidescan, vrml)
  - A **VENUS Guide** for data preservation and documentation



http://archaeologydataservice.ac.uk/research/venus.xhtml

A Brief History of the G2GPs

Digital Antiquity project to revise and expand the Guides

2009-11

Aerial Survey
Geophysics
GIS
CAD

Revised and Updated Old Guides

Excavation & Fieldwork

Retired / 'recycled'

Marine Remote Sensing
Laser Scanning
Photogrammetry

New Guides

A Brief History of the G2GPs

Digital Antiquity project to revise and expand the Guides

2009-11

- Archival strategies
- Selection and Retention
- Preservation Intervention Points
- 'Big Data'
- Creating Datasets
- Copyright

Create Integrated Archive & Project Level Sections

- Documents and Texts
- Databases and Spreadsheets
- Raster and Vector Images
- Digital Video and Digital Audio

New 'Common Components'

A Brief History of the G2GPs

Most recent development: 2013-17: **ARIADNE project**

- Initial phase allowed assessment of 14 European partner's guidelines and procedures
- Assessment phase developed a plan for new guides and case studies
  - Dendrochronology (DANS) - TRiDaS
  - Thermoluminescence (ATHENA)
  - **3D Models**

Other Guidance

- **Historic England**
  - Range of practical guides covering heritage projects

- **Digital Preservation Coalition**
  - Technology Watch Reports
  - Digital Preservation Handbook

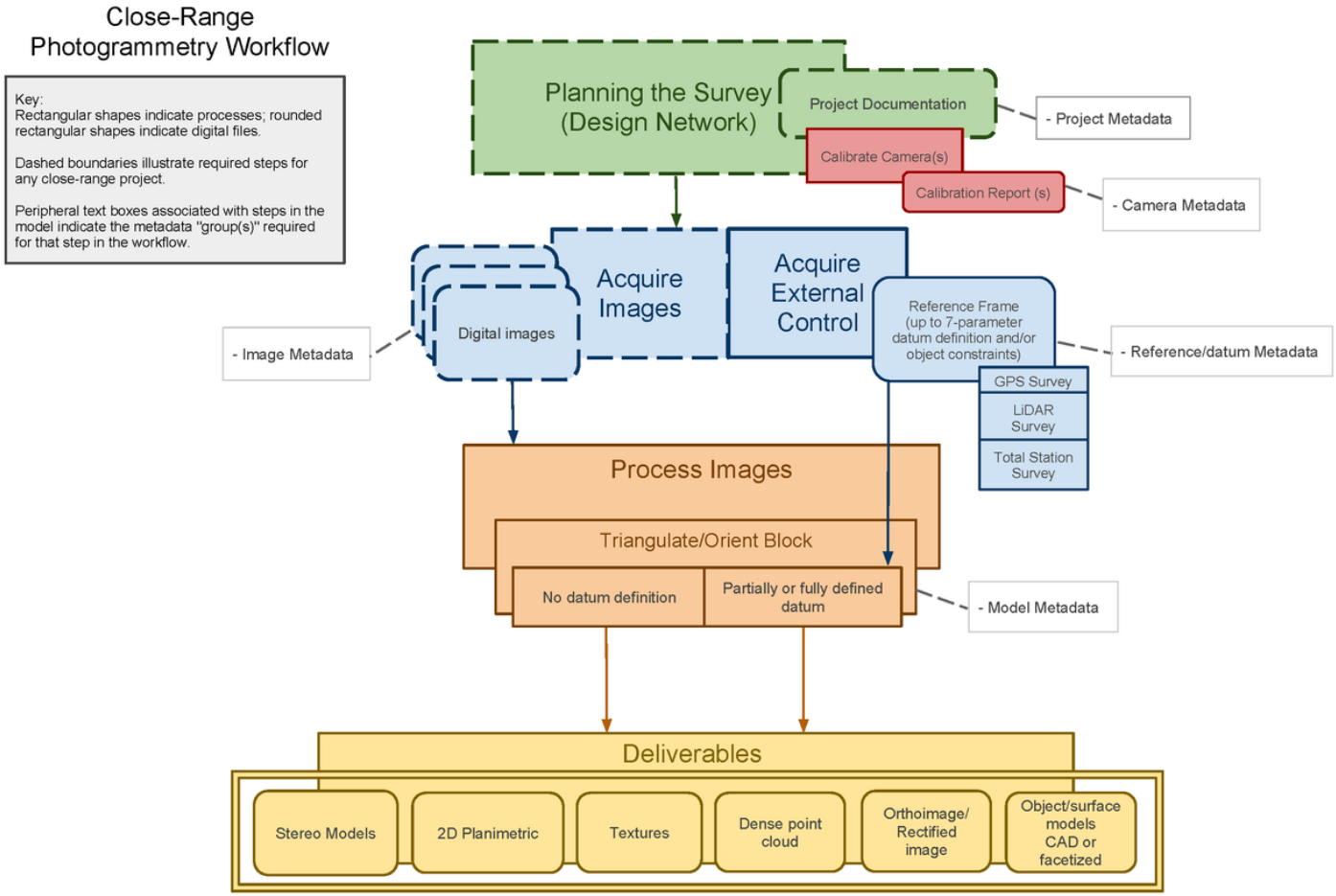
DigitalPreservationCoalition

What is 3D data?

Just models?

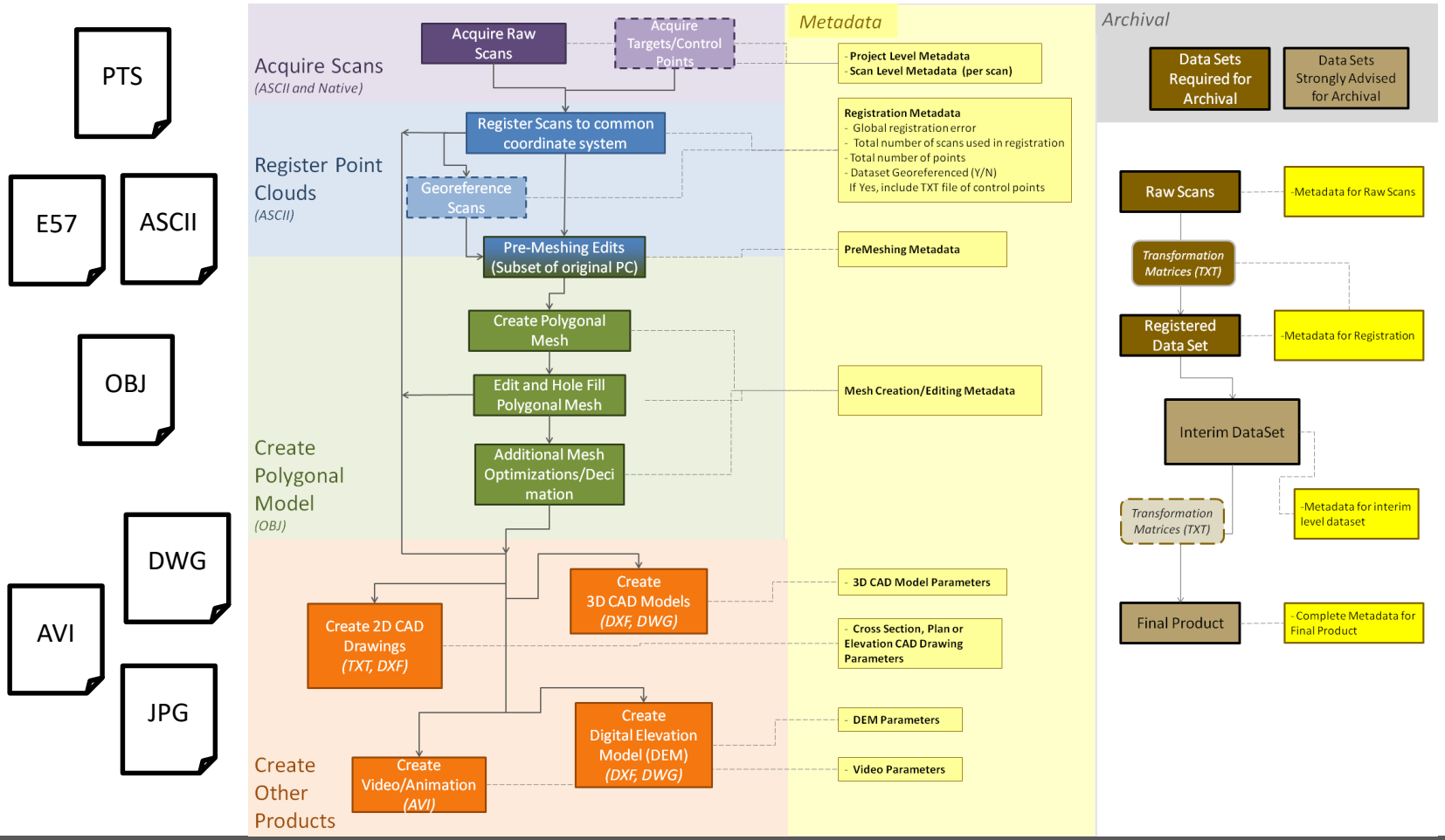
What is 3D data?

Result of different workflows and methodologies:



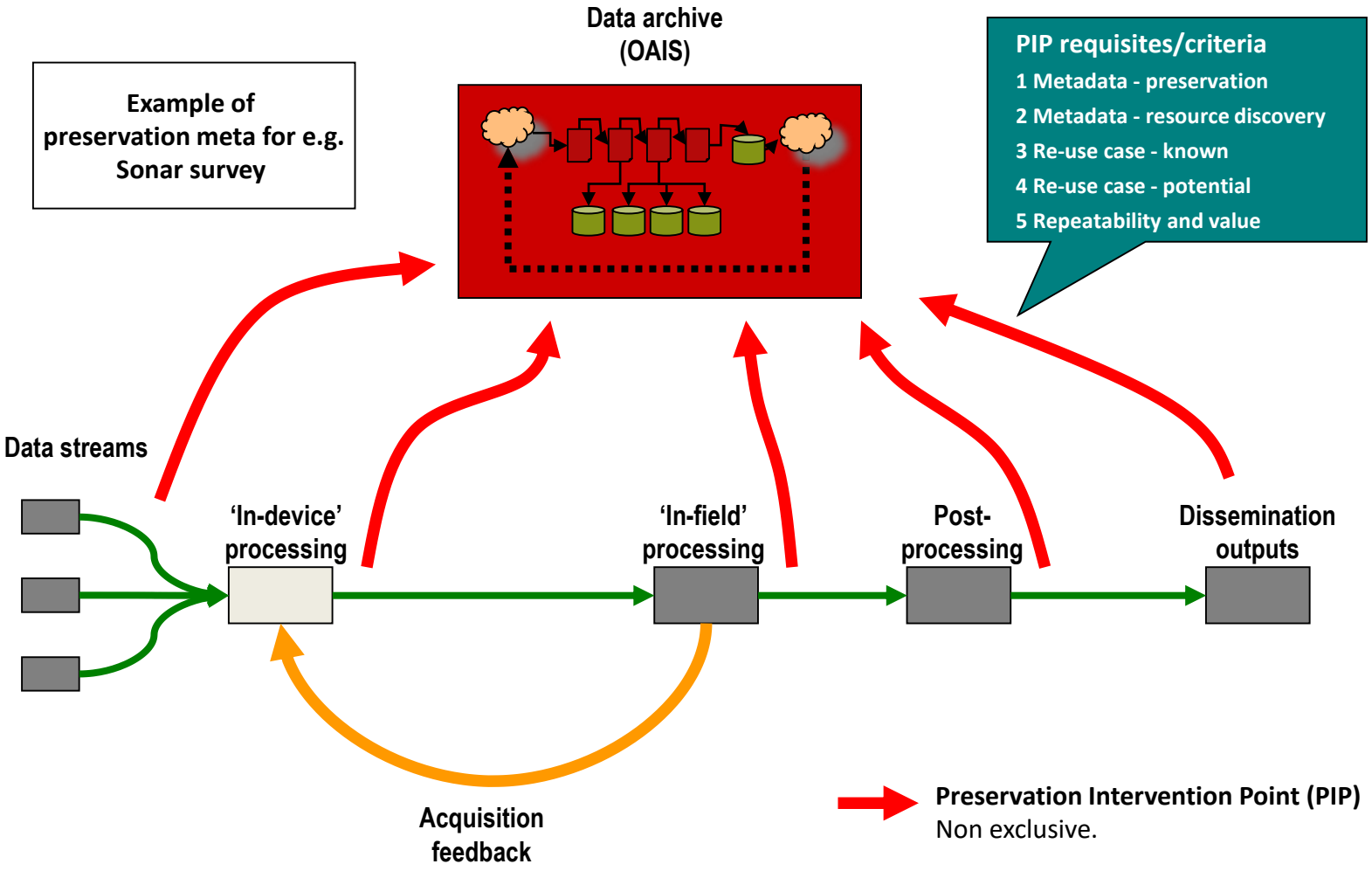
What is 3D data?

Result of different workflows and methodologies:



ads ARCHAEOLOGY DATA SERVICE

Preservation Intervention Point Schema



What to Preserve?

Assessing the workflow to:

- Identify the **raw data** for preservation
- Identify final products and deliverables - e.g. images, video, models for 3d printing
- Identify stages where data could be recreated through documentation
- Identify where data quality reduction happens e.g.:
  - e.g. RAW to TIFF to JPG
  - Images editing and masking
  - data is decimated

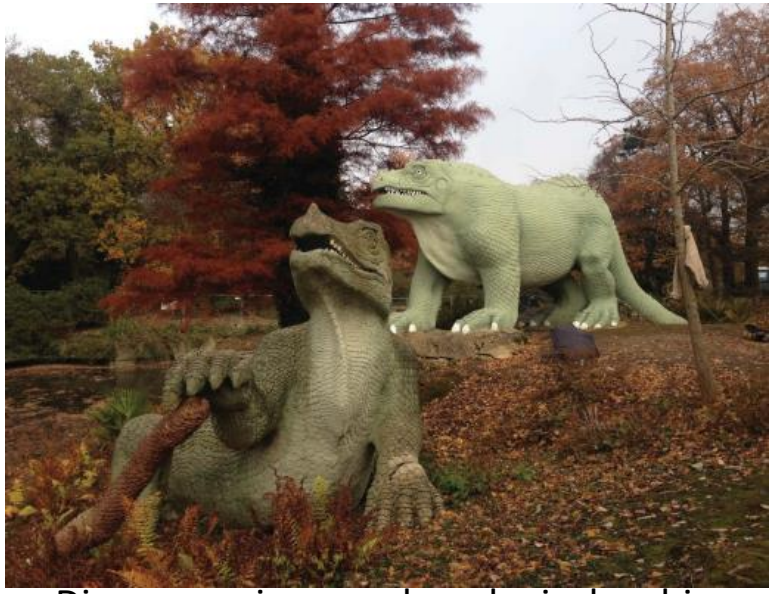
File Formats

Assessing the workflow files, at the specific points chosen:

- Are they suitable for ingest into an archive?
- Proprietary? Open? Documented? Formats
- Are they stable formats suitable for preservation?
- Are they suitable for dissemination?

Questions and Issues: Files

Some things worth considering:



Dinosaurs…in an archaeological archive

- Raw data (Pointclouds or images) can be fairly easy to preserve but can be **large**:
  - e.g. one LS collection = 3.1TB, 164k+ files
  - Crystal Palace Iguanodon survey = 15000+ images
  - Cost to the depositor – sheer numbers of files required
  - Storage is a relatively small issue but access can be problematic

Metadata

General approach:

- Metadata requirements and collection should complement the file selection process to build up a full picture of the workflow
- Record information about:
  - Data collection techniques/methodology (hardware)
  - Specific sets of files
  - Processing applied to these (in hardware and software)
- Relationships between these (CARARE2 and CRMDig)
- Technical specifications (hardware and file formats)
- Embedded metadata (EXIF, E57, etc.)

Questions and Issues: Metadata

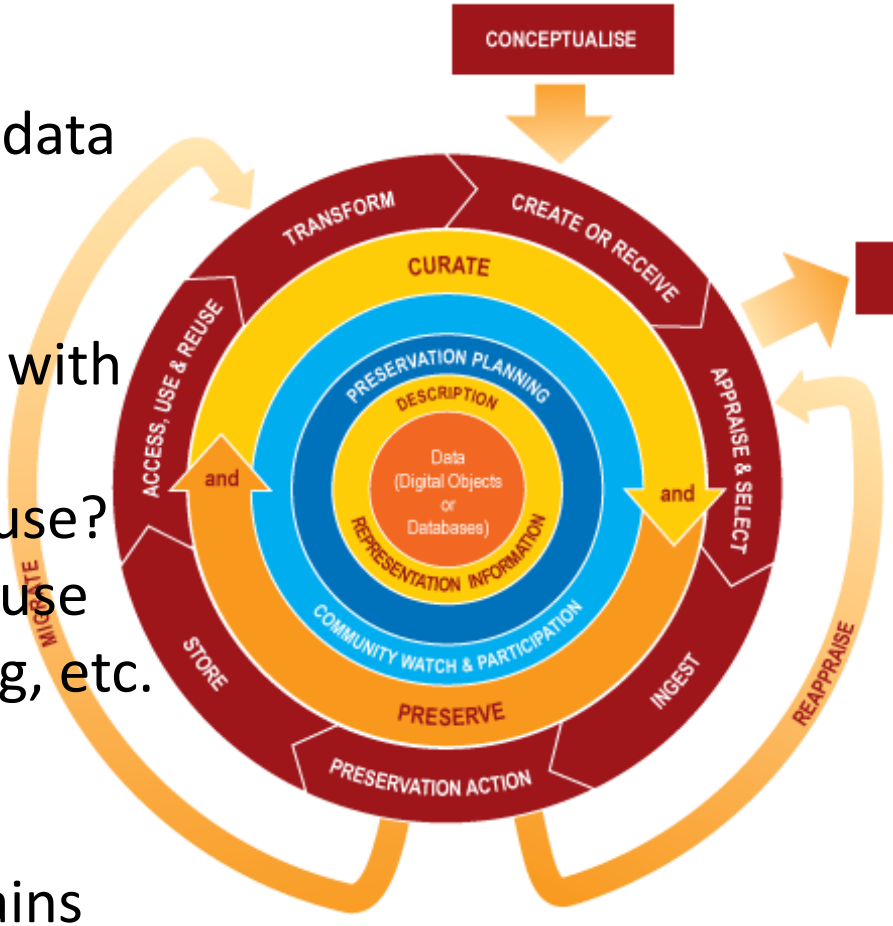
Some things worth considering:

- Workflows can be complex = metadata can be complex and lengthy
- Early engagement with creators is important:
  - Make sure the right things are recorded by those who understand them
  - Ensure that creators understand the full scope of what might be required
- Automation and extraction of embedded metadata would be ideal…
- …but we still need to define these elements and schema.

Reuse

Data reuse is key

- Reuse should, to a degree, inform data selection and documentation
- Who is reusing 3D data?
- Data creators: are already familiar with their data
- Are certain formats a barrier to reuse?
- How is it being reused? Identify reuse cases: monitoring, BIM, 3D printing, etc.
- Inappropriate reuse:
  - Survey data and security
  - Ethical issues and human remains



Preservation policy should not be a barrier to data deposition or reuse