Data Management Planning

Holly Wright
Overview

• What is data management planning?
• Why does it matter?
• What to include in a Data Management Plan (DMP)
• Intro to the DCC online Data Management Plan (DMP) system
• Group exercise
  – Drafting a Data Management Plan
What is data management planning?

Data management plans are written at the start of a project to define:

• The project
• What data will be created or collected
• How the data will be documented and described
• Where it will be stored
• Who is responsible for data security and backups
• Which data will be shared and/or preserved
• How and with whom the data will be shared
• May have preliminary, interim and final planning phases
Why does it matter?

• Advantages for research teams
  – Data is well organised, documented and in the correct formats
  – There’s no need to re-format, re-organise or try to remember details of the data
  – It’s easier to explain to new members of the team what work has been done

• Advantages for organisations
  – Uniform approach to data management by different teams
  – Clarity about which data was produced by whom
  – Researchers are responsible for managing their data well from the start
What to include in a DMP

• Description of the project
• How the data was created or collected
  – Tools and methods
  – Standards and formats
  – Ethics and Intellectual Property Rights or restrictions
• Data documentation
• Storage
  – Short term management and backups
  – Longer term archiving and preservation
• Access
  – Plan for data sharing or dissemination
What data will you create?

D. Robinson

Map of the Tropa Survey

Survey Fields

Tropa

F. Sturt Cultured Rainforest Project

Noel Tan

Lloyd-Smith

Fitjohn & Ayala

Lloyd-Smith

Cultured Rainforest Project

F. Sturt

Cultured Rainforest Project
Possible types of data

Can be anything created or manipulated on a computer:

- Text files
- Images – from digital scans of physical objects to photos
- 3D models
- Audio
- Video
- Spreadsheets & databases – numerical and textual data
- Survey data – from simple EDM surveys to Lidar scans and geophysical surveys
- Websites – even social media can be research data
- Etc...
Creating data

• What data will be created and how:
  – Are there any standards to follow?
  – Tools and software used for capturing and processing data
  – File formats – choose carefully some are better than others for long term preservation and use
  – Procedures for consistency and data quality
  – Ontologies, thesauri or controlled vocabularies used in creating metadata

• Any existing data collected or re-used

• Describe the nature, scale and scope of the data
Creating data

Guides to Good Practice

- Digital Data (general)
- GIS
- CAD
- Geophysics
- Aerial Photography
- Remote Sensing
- Fieldwork
- Virtual Reality

http://guides.archaeologydataservice.ac.uk/
Describing and documenting the data

• Data description (metadata) is essential for the future
  – A form of communication between the principle investigator and researchers re-using the data

• What metadata are needed?
  – Any standards for data archiving, data discovery or sharing to follow? Any controlled vocabularies?
  – What tools will be used for capturing metadata (Lab notebooks, Field recording sheet, Auto-saved files on instruments, Database, mobile application, etc.)
  – Formats
  – Procedures for consistency and retrievability
Storing data during the project

• Where will you store the data in the short term?
  – Local PC, network, etc.

• Describe how it is organised:
  – Project and data identifiers
  – Folder structures
  – File naming conventions
  – File version control
Data security and backup

• Identify who is responsible for data security and protection in your data management plan
• Describe your back up procedures
• How will you monitor compliance with the data management plan?
Archiving

• Plan ahead in the early stages of the project
• Identify possible archives and get in contact
• Do they have any requirements? Formats, metadata...
• Archives need permission from the owners of the IPR to preserve and distribute the data; most do not ask for a transfer of rights
Preserving for the long term

• Which data will be preserved?
  – Does the data contain any confidential personal information or high security data?
  – Are there any restrictions on access? For example
    • Privacy or ethical issues
    • Embargos for political, commercial or research reasons?
  – If there are restrictions on access, what is required to make the data available to others? Any access policies?
Access and sharing your data

• How will your data be made available for re-use?
  – Will the archive or repository disseminate your data?
  – Self dissemination through a web-site?

• How will users cite your data?

• Are there any restrictions on access?
  – License for permitted uses: non-commercial, derivatives, etc.
  – Embargo until specified date
  – Contact to request permission
• The Digital Curation Centre has a wide variety of resources online to help:
  – Checklists
  – Examples of Data Management plans from successfully funded proposals
  – An online data management planning form (DMP)
Acknowlegdements

ARIADNE is a project funded by the European Commission under the Community’s Seventh Framework Programme, contract no. FP7-INFRASTRUCTURES-2012-1-313193.

Contributors: Ulf Jakobsson, Swedish Data Archive; Hella Hollander KNAW-DANS

Image credits
Slide 6 Anti-clockwise from the top left image:
David Robinson: Chumash pictograph, South Central California, USA.
Matthew Fitzjohn and Gianna Ayala: Map of Torina survey., Italy.
Neol Tan: Digital photography at Angkor Wat. Cambodia.
Fraser Sturt: 3-D reconstruction.
Cultured Rainforest Project: Excavation at Ruma’ Ma’ on Dakah, Kelabit Highlands, Sarawak. Photograph: B. Nyiri
Lindsay Lloyd-Smith: Henge-form at Old Wolverton quarry, Milton Keynes, England. Excavation by Cambridge Archaeology Unit.