

UNIVERSITY GUIDE TO RESEARCH DATA MANAGEMENT

IMPORTANT NOTICE

During the Academic Year 2011-2012, the University will be revising its internal structures which, amongst other things, may result in the replacement during this period of the present Faculty-based structure. These changes will not amend the policies and regulations set out in this document but may result in a re-allocation of some or all of the powers and responsibilities assigned, at the time of publication, to certain Faculty officers. For this reason, the qualification '(or equivalent)' is used, as appropriate, throughout this document in relation to certain titles. Information about structural changes and any consequential re-allocations of powers and responsibilities which are agreed during the year will be published on the corporate governance web site (<http://www.herts.ac.uk/about-us/corporate-governance/home.cfm>).

Professor Q A McKellar,
Vice-Chancellor
1 September 2011

SUMMARY OF PRINCIPAL CHANGES

General changes

Refer to Vice-Chancellor's notice (above)

Refer to version 03.1, UPR IM12, for the revisions to this version.

1 INTRODUCTION

1.1 This Guide to Research Data Management was approved by the Chief Executive's Group on 6 June 2011 and sets out the arrangements for meeting the University's data management requirements set out in UPR IM12¹ and for completing the data planning elements of funding bids.

1.2 Why manage research data?

"To maintain research integrity (and protect their reputations) institutions and researchers must ensure research data remain accurate, authentic, reliable and complete for many years so that research results can be replicated by others. This requires good research data management"

(JISC: <http://www.jisc.ac.uk/supportingyourinstitution/cribsheets/datamanagement.aspx>).

By managing their data researchers and others with responsibility for research data management will:

- a meet funding body grant requirements;
- b ensure research integrity and replication;
- c ensure research data and records are accurate, complete, authentic and reliable;
- d increase research efficiency;
- e save time and resources in the long run;
- f enhance data security and minimise the risk of data loss;
- g prevent duplication of effort by enabling others to use the data;
- i comply with practices in industry and commerce.

2 CONTEXT

2.1 These guidelines should be read in conjunction with section 7 ('Research Data'), UPR IM12¹ and with UPR RE02² ('Research Misconduct') which defines research misconduct as (amongst other things) the fabrication, falsification, corruption of and/or failure to preserve research data.

¹ UPR IM12 'Data Management Policy'

² UPR RE02 'Research Misconduct'

- 2.2 Where access to data is granted to any third party or where data are routinely shared between the University and any third party, reference must be made to the Data Access Sharing Agreement (Appendix IV, UPR IM12³, refers).
- 2.3 Reference must also be made to the University's Data Security Guidance (Appendix II, UPR IM12⁴), which applies to the processing, creation, use, disclosure, dissemination and storage of person-identifiable data and other confidential and commercially sensitive data and documents (termed Personal and Confidential Data (PCD)). Appendix II, UPR IM12⁴ sets out how PCD can process safely. The guidance set out in this document (Appendix III, UPR IM12), supports the policies and principles set out in the University's Information Management Policy (UPR IM02⁵), the Information Security Policy (UPR IM03⁶) and the Records Management Standards set out in UPR IM11⁷.

2.4 External funding requirements

These Guidelines (Appendix III, UPR IM12) will help answer questions set out by external funding bodies relating to management and the sharing of data.

2.5 Consultation process

It is the responsibility of the Principal Investigator, as Data Steward, to consult at an early stage with the Research Grants Team, Business and Research Office, about funding bids and with their designated Knowledge and Business Intelligence Consultant (KBIC) in Information Hertfordshire about their data management plan, including any technical requirements.

3 DEFINITION

Research data, unlike other types of information, are collected, observed, or created, for purposes of analysis to produce original research results.

3.1 Classification of research data

3.1.1 "Research data can be generated for different purposes and through different processes (Research Information Network classification 2007):

- i **Observational:** data captured in real-time, usually irreplaceable. For example, sensor data, survey data, sample data, neuroimages.
- ii **Experimental:** data from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data.
- iii **Simulation:** data generated from test models where model and metadata are more important than output data. For example, climate models, economic models.
- iv **Derived or compiled:** data is reproducible but expensive. For example, text and data mining, compiled database, 3D models.
- v **Reference or canonical:** a (static or organic) conglomeration or collection of smaller (peer reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals."

³ Appendix IV, UPR IM12 'Data Management Policy – Data Access Sharing Agreement'

⁴ UPR IM12, Appendix II 'Data Security Guidance'

⁵ UPR IM02 'Information Management Policy'

⁶ UPR IM03 'Information Security Policy'

⁷ UPR IM11 'Records Management and the Archiving and retention of Prime Documents and Business Records'

4 DATA MANAGEMENT PLAN

4.1 A Data Management Plan ensures that all aspects of data management are fully perceived at the start of a project. A Data Management Plan is a document which describes:

- a what research data will be created;
- b what policies (funding, institutional, and legal) apply to the data;
- c who will own and have access to the data;
- d who will be responsible for each aspect of the Plan;
- e how its re-use will be enabled and long-term preservation ensured after the original research is completed;
- f what data management practices (backups, storage, access control, archiving) will be used and, therefore, what facilities and equipment will be required.

4.2 The Data Management Plan must be maintained continuously and kept up-to-date throughout the course of research.

5 DATA PLANNING CHECKLIST

The first step in data management involves producing a Data Management Plan for the research data needs of the research. Researchers are advised to use the data planning checklist in this section (5) before embarking on a project and to work with their KBIC in drawing up the Data Management Plan.

Data Description
What type of data will the research project produce? How are the data going to be used?
Roles and responsibilities
Who is the Principal Investigator (PI) who, as Data Steward, will control the data for this project? Who will be the Data Expert supporting the PI, for example, Yourself, Research Assistant, PhD Student? Is there sufficient expertise and resource in the research team to manage and store the data effectively? If additional specialist expertise is required how will this be sourced?
Standards and quality assurance
What file formats will be used? Are they long-lived? How will the (technical) quality of the data be assured? What directory and file naming convention will be used (normally University standard naming conventions are to be used)? What will be the storage and back-up strategy? Is there a research discipline standard for data sharing/integration? Is there good project and data documentation? What project and data identifiers will be assigned?

<p>Access, usage and credit</p> <p>Who will be the audience for your data and how will they use it now, and in the long run? Are there any sharing requirements, for example, a funding body's data sharing policy? Are there any other funding body requirements? Who owns the copyright and intellectual property rights of the data? What kind of possession issues are involved? Who will decide on access to the data? What kind of rights will be granted to different user groups for reading and managing data files? How will confidentiality be ensured? How will data protection be carried out? Are there any special privacy or security requirements, for example, personal data, high-security data? When and where the data will be published? Will data be deposited in appropriate repositories in a timely manner?</p>
<p>Benefits and cost effectiveness</p> <p>Are there any costs associated with the data management and sharing plan? How much data will be generated and how often will it change? Are there tools or software needed to create/process/visualise the data? Are there appropriate computing hardware, facilities and resources to manage, store and analyse the data?</p>
<p>The research process</p> <p>How will the data be collected? How will the research participants be informed? Which software will be used in storing and processing the data? Which data formats and storage media will be used? How will data processing be documented?</p>
<p>Preservation and sustainability</p> <p>What type of data will be produced? Will it be reproducible? What would happen if it got lost or became unusable later? What kind of data will be collected? What kind of data and file backup procedures will be used? What will happen to the data after the original research is completed? How will the metadata on the data collection and dataset content be stored? How long should it be retained, for example, 5 years, up to 10 years, or permanently?</p>

P E Waters
Secretary and Registrar
Signed: 1 September 2011