

The Data Wave



“Riding The Wave” European Commission October 2010

Image shown is from front cover of the 2010
report of the High Level Expert Group on
Scientific Data

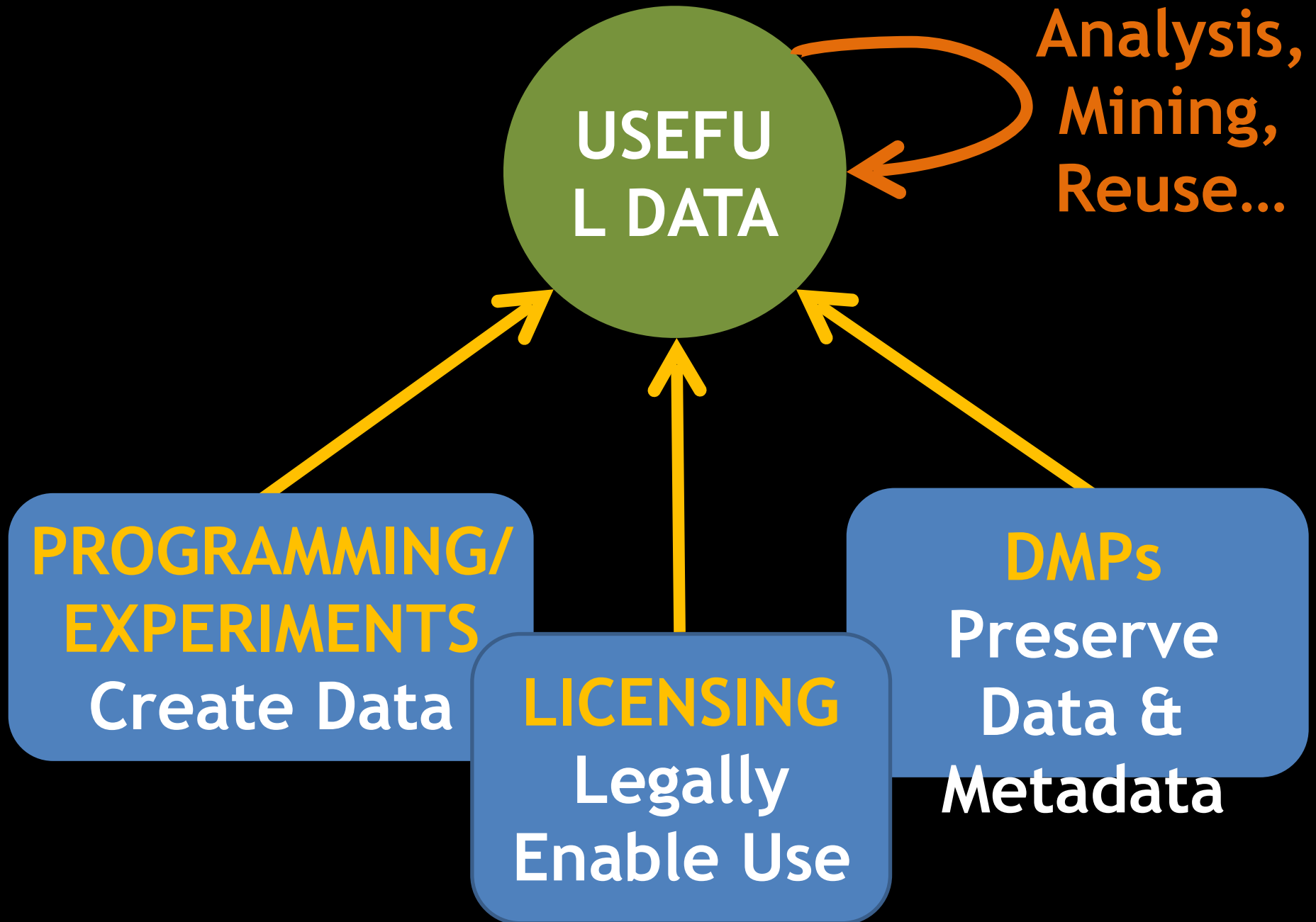


“In one day, a high-throughput DNA sequencing machine can read about 26 billion characters of the human genetic code. That translates into 9 terabytes – or 9 trillion data units – in the course of one year...”

Image shown is from front cover of the 2010 report of the High Level Expert Group on Scientific Data

So why is this relevant to me?

- ① **Medical Image Analysis:** May account for 30% of all data storage
- ② **Astronomy:** Square Kilometre Array, 1 petabyte of data every 20 seconds?
- ③ **Systems Biology:** Genes, proteins, GenBank, “1000 Genomes”...
- ④ **Computationalists:** Large-scale simulation data, HPC...



What's In a DMP?



RESEARCH PHASE

- file formats
- ownership
- metadata
- storage
- backups

DISSEMINATION PHASE

- share with whom?
- release date?
- licensing
- metadata

PRESERVATION PHASE

- repository?
- long-term manager?



RESEARCH
PHASE

DISSEMINATION
PHASE

PRESERVATION
PHASE

A DMP:

- helps you to **STRUCTURE** your research output;
- improves the **EFFICIENCY** of your work;
- increase **LONGEVITY** of your research;
- enables **LARGE-SCALE** exploration in

Digital Curation Centre

How to Develop a Data Management and Sharing Plan

Introduction

This guide outlines the process of developing a data management and sharing plan. Planning for the effective creation, management and sharing of your data enables you to get the most out of your research. The guide outlines UK funder expectations for data management and sharing plans and provides practical guidance on how to meet these. It should be of relevance to researchers and those supporting them within higher education.

Why develop a data plan?

There are many benefits to managing and sharing your data:

- you can find and understand your data when you need to use it
- there is continuity if project staff leave or new researchers join
- you can avoid unnecessary duplication e.g. re-collecting or re-working data
- the data underlying publications are maintained, allowing for validation of results
- data sharing leads to more collaboration and advances research
- your research is more visible and has greater impact
- other researchers can cite your data so you gain credit

What do research funders want?

Many UK funders have released data policies which advocate curation and data sharing¹. Several of these require that data management and sharing plans are submitted as part of grant applications². Funders expect data plans to outline how data will be created, managed, shared and preserved, justifying any restrictions that need to be applied. The plans are an opportunity to demonstrate your awareness of good practice and reassure funders that your proposal is in line with their data policy.

The DCC has collated UK funders' guidelines on what to cover in data management and sharing plans³. Funders typically propose broad themes or questions for you to consider as appropriate to your proposal, through the Arts and Humanities Research Council (AHRC) and Economic and Social Research Council (ESRC) ask you questions.

DMPs in this assessment

DMPs (Phase 1 only)

- ① Follow the email link to your DMP questionnaire**
- ② Answer the 20 Questions in relation to your Phase 1 work**
- ③ Save the DMP as an .xml file and include a copy in your repository**
- ④ Only ONE DMP to be made per group**

Useful Tools and Sites

Digital Curation Centre (@digitalcuration on Twitter)

Provides advice on data management and collaborates on many data curation projects

- Online DMP tool (requires registration and login)
- Advice leaflets and data curation guides
- “How to Develop a Data Management and Sharing Plan” (see Weblearn)

Oxford-Specific Resources

Leaflet “Managing Your Research Data at the University of Oxford” (see Weblearn)

HFS Services: <http://www.oucs.ox.ac.uk/hfs/index.xml>

Blogs

Peter Murray-Rust: “Why you need a data management plan” (plus a wealth of other posts on Open Science and Open Data) <http://blogs.ch.cam.ac.uk/pmr/category/data/>

Data Resources

Lists of Open Definition conformant licenses: www.opendefinition.org