Research data curation – the funders' perspective

Kevin Ashley
Digital Curation Centre
www.dcc.ac.uk
@kevingashley
Kevin.ashley@ed.ac.uk

The DCC is supported by Jisc

Background

- UK research funders have requirements about preservation of data from funded research
- Universities, researchers unclear about who pays to meet these requirements
- DCC has a national coordinating role, so…
- We got everyone together in one room to talk about it
- This is what we found….

An aside

- This work from the 4C project:
  - "**C**ollaboration to **C**larify the **C**osts of **C**uration"
- One goal is to collect real data on curation/preservation costs: www.4cproject.eu
- We have many models. This is not another.

Collaboration to Clarify the Costs of Curation

It's not about costs

- Very few are asking 'what does this cost' – they think they know
- It is about 'who pays' or 'how do I account for these costs'
- There is actually a strong relationship between these two questions
  - This is not specific to research data

Requirements & reasons

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

See: http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

The 1-minute version

- Research data are a public good – make openly available in timely & responsible way
- Have policies & plans. Data with long-term value should be preserved & usable
- Metadata for discovery & reuse. Link publications & data
- Sometimes law, ethics get in the way. We understand.
- Limited embargos OK. Recognition is important – always cite data sources
- OK to use public money to do this. Do it efficiently.

The UK funding model

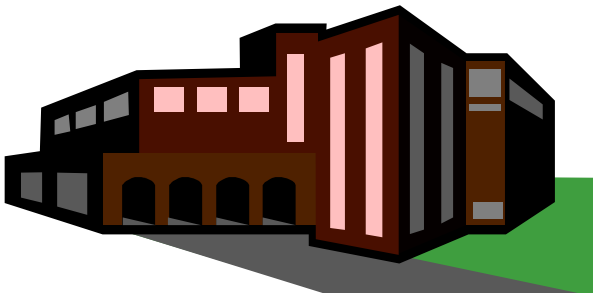
Research Council

Project funding, to PI

Higher Education Funding Council

Annual funding, per institution

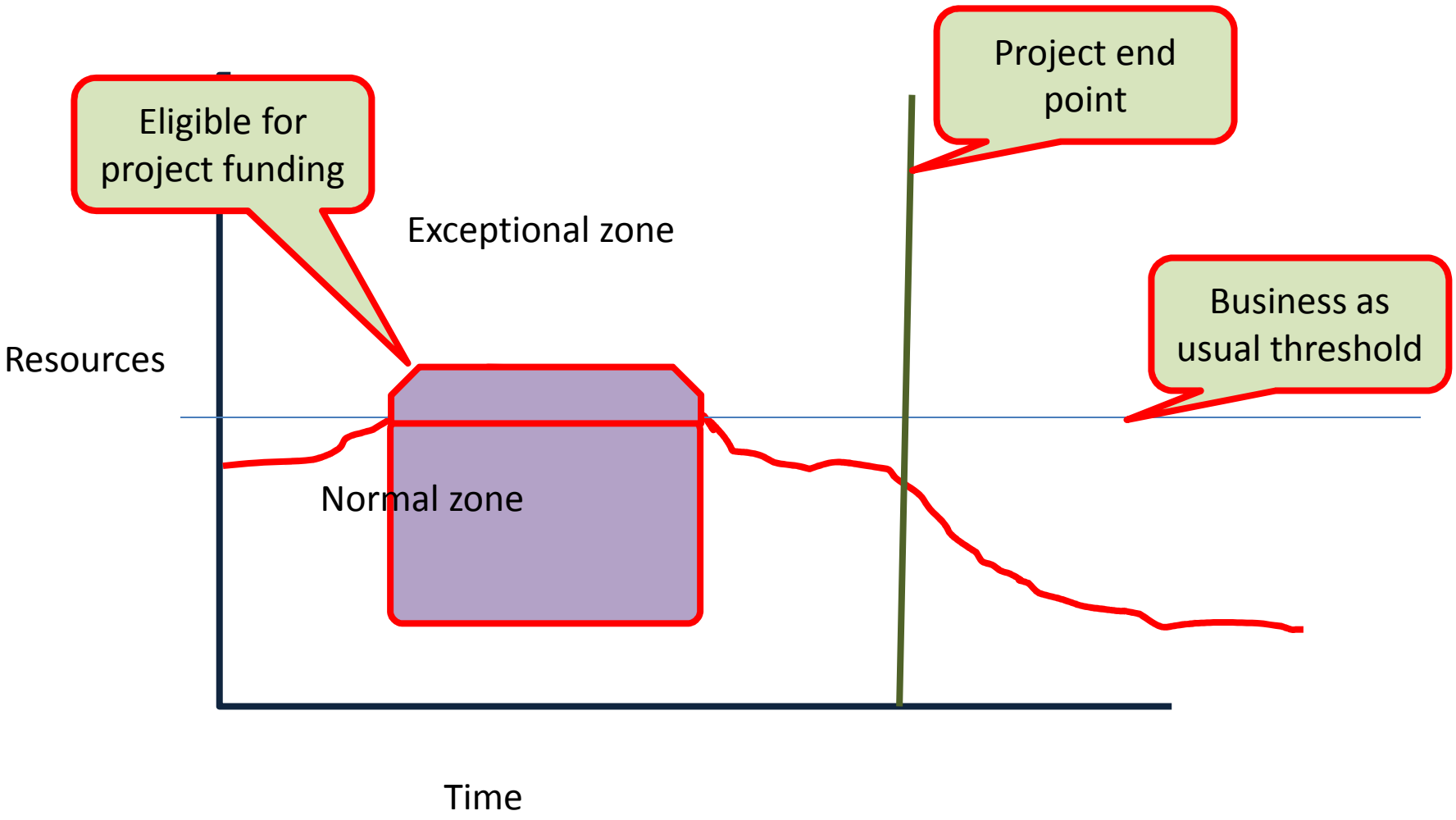
The complexities appear in every country, just in different ways



What it means

- Project funding can only be spent during projects on direct project costs
- Project funding comes with overheads, which universities must use for research infrastructure
- Ongoing ('QR') money is continuous, relates to research ranking
- Important to distinguish business-as-usual from exceptional requirements

A research lifecycle



We have money

Funders view

We have requirements

We have rules about how you use money to meet requirements

Over to you!

Being clever with costs

- Ongoing costs beyond project end cannot be charged to a grant, but…
- ‘Pay once, store forever’ charges acceptable.
- Thus, incentive to outsource lon curation
- Yet universities are only acting as last-resort option in any case – discipline data archives preferred

Many of these are run by funders

Controlling costs

- Not everything to be kept, and not all forever
- 10-year minimum for data ‘of value’
- Who makes retention decisions?

Observations

- Funders not necessarily interested in costs unless it is for infrastructure they operate
- Fixed costs for infrastructure more significant than marginal costs per project
- Funding rules can make outsourcing attractive (but can increase costs as a result)
- The system does not always operate to minimise total cost

Models & players

- Services geared to ‘data publishing’ workflow
  - Dryad – data behind publication
  - Arkivum – and similar commercial players
  - Figshare – targeting the researcher (and now institution)
  - EUDAT – copy existing model, invent a new one?
- One off payments from project costs, or…
- Institutional subscription (from overheads?)
- Active data only handled by some data centres

Questions

- Target researcher or institution?
- How much commonality across Europe?
- Do different disciplines require different models for funding?
- Can EUDAT offer efficiencies or reduce financial risk?