



# RDM Service Models

## Research Data Management

Thomas Parsons, Laurian Williamson  
February 2013

## Assumptions

- The activities identified in the document are a complete summary of funders' requirements
- The activities, leadership and support roles identified in the document are acceptable as a basis for the service

## Requested by steering group

- Minimal model to match compliance requirements and pragmatic models for delivery:
  - Minimal
  - Mediated
  - Consultancy

## Overview

- Definition of service levels - slide 4
- Benefits & risks - slides 5-7
- Current status – slide 8-9
- Service model timelines – slides 10-12
- Ownership – slide 13
- Resourcing – slide 14

<b>Core service activities</b>	<b>Level 1 Minimal</b>	<b>Level 2 Mediated</b>	<b>Level 3 Consultancy</b>
Data Management Plans & advice on funding body requirements	PI with website guidance	Training, advocacy & some one-to-one	Tailored approach, subject-specific advice & training
Active data management & storage	IT Support, Schools & PI (ad hoc)	Training, advocacy & some one-to-one	Tailored approach & subject-specific advice
Data archive & preservation	IT Services with IT Support	Training, advocacy & some one-to-one	Tailored approach & subject-specific/funder requirements
Data sharing & publishing – advice on funding body requirements	IT Services with IT Support	Training, advocacy, CPD at UoN	Tailored approach & on-going training
Copyright & IPR	PI with website guidance	Training, CPD at UoN	Tailored approach & on-going training
Compliance & reporting	PI responsible	Institutional overview & mechanisms	Systemised reporting
	<i>Website advice and self-supporting</i>	<i>Focused support and capacity building</i>	<i>Consultancy, subject-specific and embedding</i>

## Benefits & risks – Level 1 Minimal

- Benefits:
  - No additional resource required
  - Engages IT Support strongly with the research community
- Risks:
  - PI led, so assumes a level of competence & understanding
  - Transfers administrative tasks to PI, reducing time for research
  - PI has to find multiple sources of information & people
  - No provision for awareness and advocacy
  - RDM website will need regular updating reflecting sector change
  - All reporting & compliance is reliant upon PIs
  - No capacity building within Schools

## Benefits & risks – Level 2 Mediated

- Benefits:
  - Systematic advocacy, training & awareness sessions
  - RDM website managed and updated
  - Targeted RDM services for RCUK & research priorities
  - Embedded subject-specific expertise increases over time
  - Reduces time taken for DMPs & improves quality
  - Team facilitates reporting & compliance
  - Dedicated staff provide RDM landscape awareness
- Risks:
  - Mediated service, demand may exceed capacity (e.g. DMPs)
  - Staff up-skilling required prior to service launch
  - Additional resource will be required (2.0 FTEs minimum)

## Benefits & risks – Level 3 Consultancy

- Benefits:
  - All benefits from Level 2
  - Capacity building and dedicated Faculty representatives
  - High degree of subject-specific advice
  - RDM website would include tailored advice
  - Clear overview of research projects & compliance to mandates
  - Improvement in quality of DMPs and proposals
- Risks:
  - Recruiting subject-specific RDM expertise
  - Requires embedding at Faculty or School level (but may already exist in some areas e.g. Life sciences)

## As it stands

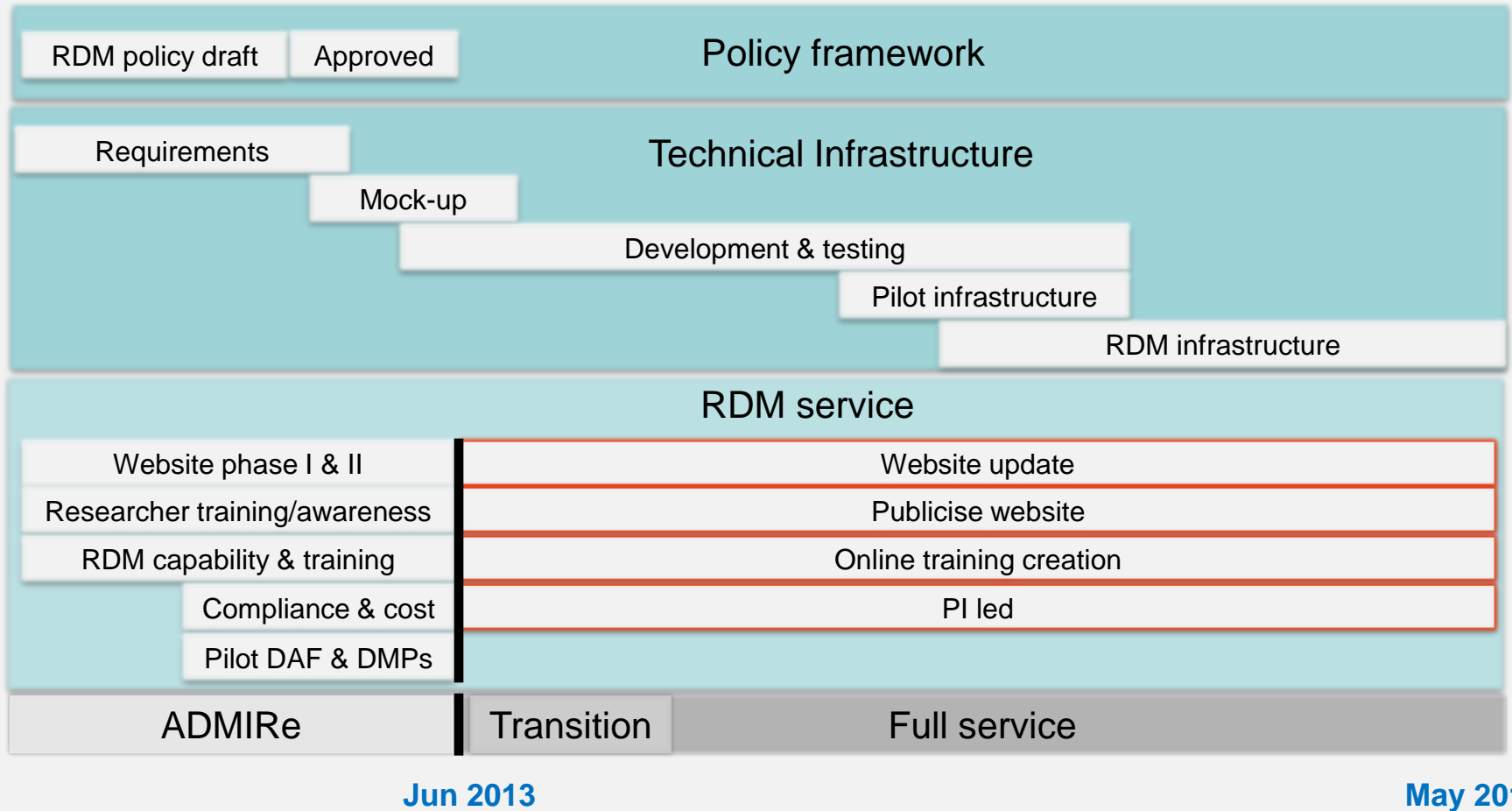
- Funders' and institutional requirements are not being met with the current provision (25% write DMPs)
- Level 1, 2 or 3 service would meet funders' requirements with various levels of quality and future-proofing
- Researchers' expectations indicate a desire for training and support beyond website advice and self-supporting
- Competitor institutions have investment in this area:
  - Edinburgh (£1million & £250k per year)
  - Bristol (7 FTEs proposed)
  - Manchester (5 FTEs in place)



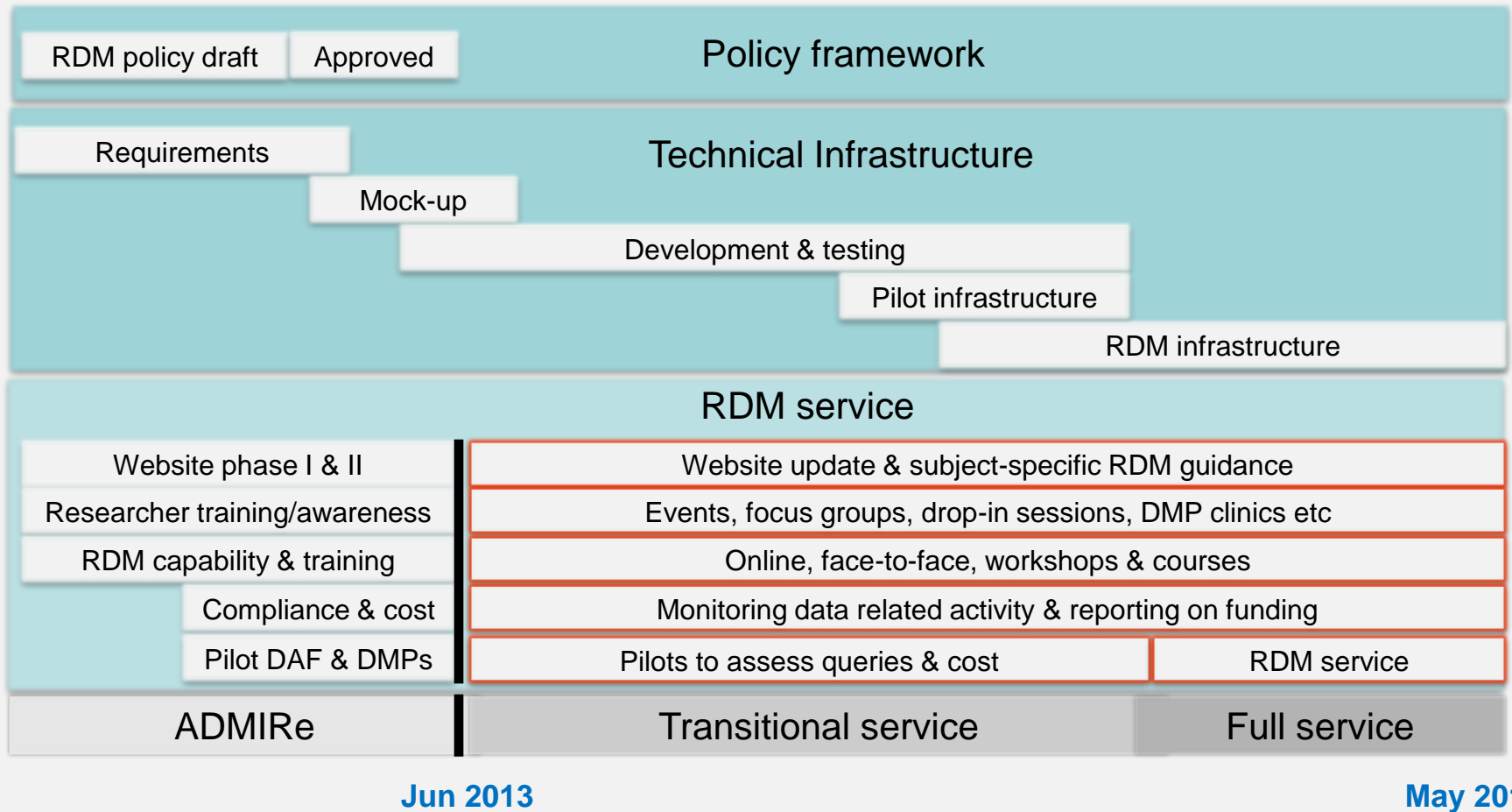
## Moving forward

- Pre-requisites for any service:
  - An RDM policy
  - Development of the technical RDM infrastructure (IT Services)
  - Transitional RDM support service from ADMIRe to long-term
- ADMIRe is delivering **elements** for a Level 2 service:
  - Researcher engagement (DMPs, pilots, survey, promotion)
  - Technical & preservation requirements
  - RDM website
  - Advocacy, events & awareness
  - Training for staff & researchers

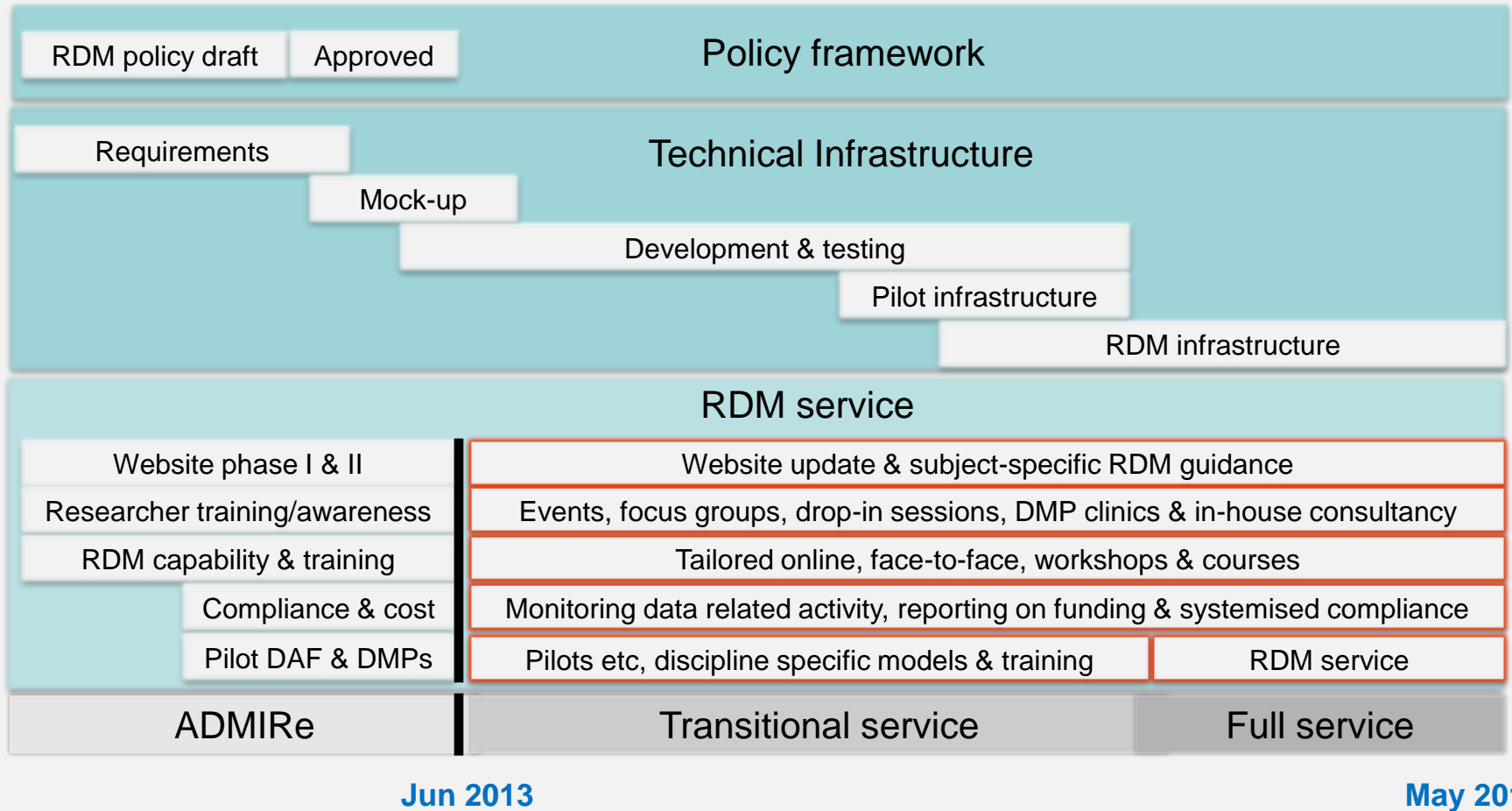
# RDM service timeline Level 1



# RDM service timeline Level 2



# RDM service timeline Level 3



Jun 2013

May 2015

## RDM process ownership:

- **RGS**
  - Policy perspective and driven by compliance
  - Financial focus on the proposal stage
- **IT Support**
  - Already carrying out components of RDM e.g. storage
  - Recognised as experts in certain areas e.g. data security, DMPs technical appendices
- **LRLR**
  - Recognised RDM expertise & activities (ADMIRe)
  - Strengths in OA & research communications
  - Metadata expertise
- **Distributed – PI or School as owners**
  - Pockets of activity & high subject expertise
  - Low awareness of RDM as a whole
- **No department has engaged systematically**

## Resourcing

- Level 1
  - Process and website adopted and promoted
  - Support staff up-skilling
- Level 2
  - 2.0 FTEs additional minimum
  - Researcher training & awareness
- Level 3
  - 3.5 FTEs additional minimum
  - Embed within Faculty and at School level to build capacity
  - Assumption: after 2-3 years PIs will be familiar with RDM – dedicated FTEs could be reduced

## Service expectations

- Focus on:
  - Data that underpins a publication
  - RCUK projects
  - Advocacy & training
- Compliance led:
  - 629 applications for RCUK (2012)
  - 1-4 hrs per DMP (~0.5-1.5 FTE)
- Research impact led:
  - 4000 publications in 2012 (~40% RCUK) = 1600 publications per year
  - 0.5-4 hrs per data set published (~0.2-1.5 FTE)
  - Requires LRLR & BEIS expertise

## Service expectations

- Advocacy & training:
  - 0.2-0.5 FTE across all Faculties (assumes online & CPD training course)
  - Generic at start, moving to subject specific
- Ad-hoc support requests:
  - Estimate 4-10 requests per week, 0.1 to 4 hours per request (0.1-0.4 FTE)
  - Accurate data only available over time and linked to advocacy activity



## IT Support/LRLR – RDM queries (all time)

RDM activity	IT Services	LRLR
Planning a research idea:	8	3
Systematic review of the literature/data review:	1	4
Funding/Grant bid proposal:	10	2
Data management planning:	11	1
File organisation:	19	2
Local storage:	25	2
Data gathering:	9	4
Documentation and Metadata:	6	2
Data analysis:	8	0
Publishing research data:	5	1
Repository storage:	11	0
Data curation:	2	2
Data deposit:	1	0
Data discovery:	4	3
Accessing and re-using research data:	5	2
Never dealt with a RDM query:	10	10
Other (please specify):	2	1
<b>Total</b>	<b>137</b>	<b>39</b>
<b>Number of respondents</b>	<b>36</b>	<b>17</b>

# DMPs MRC example:

## 1.3 Format and scale of the data

*Clinical Measurements are maintained in a Microsoft Access for Data Files Database as Microsoft Access 2007 accdb files. This enables validity of the data but not sharing.*

*Microarray data will be acquired in the Affymetrix GeneChip Software and analysed in MATLAB and R, which will enable long term validity and sharing.*

## 3.1 Managing, storing and curating data.

*Clinical measurements - stored securely on University of Nottingham Servers at Kings Meadow Campus. Data files will be stored for the length of time denoted in each study Ethical approval, University of Nottingham SOP and/or other Legislation as required. All data is stored on University Servers is backed up by University of Nottingham I.T. Services on a daily basis. A local back up of all NRBRU data files is also carried out on a daily basis. Local Back-Ups are performed using "ACEBIT ACEBACKUP v3" software. Local backups are ENCRYPTED using AES256 bit encryption and backed up to an external hard drive located in Room B28 of the Clinical Sciences Building, City Hospital, Nottingham. Email notification with full report is sent to Database Manager Glenn Hearson nightly on files backed up and of any errors encountered. A total of 5 local Back Ups are carried out each week, 1 for each day of the working week, i.e. Monday to Friday inclusive. Back-ups are performed each evening at 23:45pm. Each local back up is retained for a period of 7 Days.*

*Microarray data - will initially be collected at the University of British Columbia in Dr Kobor's lab. The data is collected on a dedicated ChIP-chip computer; backed up on an external hard-drive and network server; managed by a graduate student in the lab who has expertise with ChIP-chip data and files. The data will then be transferred to the University of Nottingham via web enabled secure and backed up research "filestore" facility. Upon my return to Nottingham the data will be transferred to centrally managed, secure and backed up storage space, of which up to 8 terabytes is available to me.*

# DMPs MRC example:

## **3.3 Data preservation strategy and standards**

*Clinical measurements - Data files will be stored securely on University of Nottingham Servers at Kings Meadow Campus, for the length of time denoted in each study Ethical approval, University of Nottingham SOP and/or other Legislation as required.*

*Microarray data - microarray data will be retained for a minimum of 10 years as per the University of Nottingham Data Management Policy. Data will be uploaded to the ArrayExpress public database and also to the University of Nottingham secure, backed up and validated data repository*

## **5.1 Suitability for sharing**

*Both anonymised clinical measurements and microarray data will be available for sharing*

## **5.2 Discovery by potential users of the research data**

*The clinical measurements will only be made available via direct contact with the Nottingham research group. Information required by the user will be supplied at the same time as the cells.*

*Microarray data will be available via the ArrayExpress Experiments archive which can be "browsed" and searched. Equally the University of Nottingham Repository will be indexed with sufficient metadata to allow discovery by users. Users will gain access following approval from the PI or other named responsible person.*

## DMPs ERC example:

14,000 word document

### Overview:

This research project consists of three key methodologies: scientific analysis of glasses and glazes, genetic studies, and the combination of these data using digital techniques, geospatial science and statistics.

### Outputs:

**A semantic web based server** will be created so that the archaeological, historical and scientific information will be interrogated together comprehensively.

## DMPs ERC example:

Deliverable D3.1 **a database** and a fully tested/ validated knowledge platform and knowledge base that will allow for the input of archaeological and historical knowledge and disparate data types so as to make it possible to combine the data and carry out statistical and semantic integration and interrogation generating innovative knowledge about the Silk Road.

The dissemination of this project will be through attendance at a minimum of four national and four international conferences, the Museum installation will be displayed in at least four Museums and in the British Library, the publication of multi-authored articles in peer-reviewed international journals like Proceedings of the National Academy of Sciences, Antiquity, the Journal of Archaeological Science and Archaeometry. Funding is also included for costs of translating articles into Arabic and Chinese so that they can be distributed to scholars in the Arabic world and China.

The results will provide **a validated platform** that can be adopted by researchers in a wide range of disciplines in public institutions like Museums, Libraries and Universities world-wide.

