

Taking the analysis to sensitive data. Test cases with the British Library Digital Data and F1000 Research Data

27/2/2015

**AMASED** Access Methods for Analysing Sensitive Data  
Research data spring

- » Lead: University of Bristol
- » Collaborators: London Metropolitan University, Content Mine, British Library, F1000 Research
- » Stakeholders: research data providers, researchers, research libraries, publishers etc.

- » Sensitive data crucial in almost all areas.
  - » Patchy understanding of how to deal with them (3.9). Tools and methods evolving rapidly.
- » DataSHIELD ([www.datashield.ac.uk](http://www.datashield.ac.uk)) developed for Biomedical/Life sciences (EU,MRC)
  - » Extend to new roles in biomedicine, text data in humanities, and work with 3.3 to develop capacity for data cleaning
- » Goal 1: Scope integration of the data cleaning tool ICT-RD (3.3) with DataSHIELD
- » Goal 2: Local instance of DataSHIELD for open text data (digitised books) from the British Library
- » Goal 3: Scope the challenges of implementing DataSHIELD as a paper data access-analysis solution for F1000 Research.
- » FINAL (Month 13): Final products implemented with London Met, British Library, F1000 Research

**Ethics**

Explain DataSHIELD to ethical and data access committees

Hackathon will feed into ongoing evaluation by the proposed DataSHIELD Security Oversight Committee

**Potential developers**

Raise interest in new project to optimise the exploitation of biomedical data, opportunity to develop new functionality

**New use cases**

Raise awareness of DataSHIELD beyond scope of biomedical community. Application to other research areas?

**Researchers**

Increase their access to datasets globally  
Simplify the process of pooled data analysis  
Ability to analyse individual level data

**Data providers**

Increase usability of datasets globally  
Preserve intellectual property  
Preserve participant confidentiality

- » Open source products
- » Free to user and data provider
- » Growing international development community
  - » Protocols for quality control
- » Open for others to join us

**DELIVERABLES**

1. Identify a workflow and methodology to implement the ICT-RD data cleaning tool in existing DataSHIELD infrastructure for numerical data.
2. Locally deployed DataSHIELD test infrastructure for text analysis of digitized books
3. Explore and identify a model for an F1000 Research DataSHIELD infrastructure for analysis of data in their papers.

**MEASURES OF SUCCESS**

1. Defined and realistic methodology for integrating  
The two pieces of software
2. Achieve unrestricted textual analyses of openly available text data (digitised books) using a locally deployed DataSHIELD test infrastructure
3. Defined and realistic methodology for implementing DataSHIELD on F1000 research data

		Cost including Inflation		Cost excluding Inflation	
		YEAR 1	TOTAL	YEAR 1	TOTAL
		£	£	£	£
<b>Staff Costs</b>	<b>FTE</b>				
Dr Rebecca Wilson					
Basic Salary		0	0	0	0
National Insurance		0	0	0	0
Superannuation		0	0	0	0
Total	0.0	0	0	0	0
Grade J					
Single Spine (Grades A-M) - Grade J - Increment point 1 - Universities Superannuation Scheme					
Basic Salary		6,728	6,728	6,728	6,728
National Insurance		528	528	528	528
Superannuation		1,077	1,077	1,077	1,077
Total	0.191	8,333	8,333	8,333	8,333
Single Spine (Grades A-M) - Grade I					
Single Spine (Grades A-M) - Grade I - Increment point 2 - Universities Superannuation Scheme					
Basic Salary		1,466	1,466	1,466	1,466
National Insurance		112	112	112	112
Superannuation		235	235	235	235
Total	0.045	1,813	1,813	1,813	1,813
<b>Total Staff Costs [1]</b>		<b>10,146</b>	<b>10,146</b>	<b>10,146</b>	<b>10,146</b>
<b>Non Staff Costs</b>					
Travel & Subsistence		2,000	2,000	2,000	2,000
<b>Total Non Staff Costs [2]</b>		<b>2,000</b>	<b>2,000</b>	<b>2,000</b>	<b>2,000</b>
<b>Facility Costs</b>					
<b>Total Facility Costs [3]</b>		<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Estate Costs</b>					
Estate Costs		1,844	1,844	1,844	1,844
Infra Lab Technician Costs		0	0	0	0
<b>Total Estate Costs [4]</b>		<b>1,844</b>	<b>1,844</b>	<b>1,844</b>	<b>1,844</b>
<b>Indirect Costs</b>					
<b>Indirect Costs [5]</b>		<b>10,264</b>	<b>10,264</b>	<b>10,264</b>	<b>10,264</b>
<b>TOTAL COSTS [6=1+2+3+4+5]</b>		<b>24,254</b>	<b>24,254</b>	<b>24,254</b>	<b>24,254</b>