

Phase	Months	Goal	Objectives	Deliverables	Indicators of Success	Work Assigned	Costing
Phase 1: Scoping	month 1 - month 3	Goal 1: Scope the challenges of integration of the data cleaning tool ICT-RD (dataspring proposal 3.3) with DataSHIELD (www.datashield.ac.uk).	- Liaise with researchers at London Metropolitan University to establish a method to integrate ICT-RD data cleaning tool with DataSHIELD.	Identify a workflow and methodology to implement the ICT-RD data cleaning tool in existing DataSHIELD infrastructure for numerical data.	A defined and realistic methodology for integrating the data cleaning tool within DataSHIELD	University of Bristol, London Metropolitan University	Costing 1
Phase 1: Scoping	month 1 - month 3	Goal 2: Set up a local instance of DataSHIELD to analyse openly available text data (digitised books) from the British Library without any controls on disclosure or licensing.	- Procure a subset of openly available text data (digitised books) from the British Library - Liaise with digital humanities researchers to shortlist core analyses that are required and implement these from existing R packages. - Investigate and identify potential areas of disclosure / licensing or IP conflict using standard R analysis packages to analyse digitised book data. - Create a test infrastructure: * Install a local DataSHIELD instance. * Server side - define how raw digitised books (XML files) are stored/accessed by server side DataSHIELD. * Client side - use existing R text analysis packages to connect to server side DataSHIELD to conduct unrestricted analysis. * Authentication: Define how client side functions will authenticate with the server to satisfy the above disclosure/licensing issues.	Locally deployed DataSHIELD test infrastructure.	Achieve unrestricted textual analyses of openly available text data (digitised books) using a locally deployed DataSHIELD test infrastructure.	University of Bristol, Content Mine, British Library	
Phase 1: Scoping	month 1 - month 3	Goal 3: Scope the challenges of implementing DataSHIELD as a paper data access-analysis solution for F1000 Research.	- Liaise with F1000 Research to identify an example dataset from published papers or papers in review. - Explore models of a DataSHIELD test infrastructure that are compatible with the requirements of F1000 Research	Explore and identify a model for an F1000 Research DataSHIELD infrastructure for analysis of data in their papers.	A defined and realistic methodology for implementing DataSHIELD on F1000 Research data	University of Bristol, F1000 Research	
Phase 2: Market Research	month 4 - month 5	Goal 1: Set up an advisory group to investigate the key text analytical techniques required in DataSHIELD and to understand data restriction with respect to digitised text.	- Invite a group of digital humanities researchers, text mining experts and those involved in digital collections to join the advisory group - Investigate core analyses that are required for text analysis in digital humanities, explore existing R packages/functions containing suitable functionality - Investigate and identify potential areas of disclosure of restricted digitised book datasets - Investigate DataSHIELD methods (statistical/computational) to prevent disclosure of re	Shortlist relevant analytical techniques and R packages for text analysis. Produce workflow to prevent data disclosure, adhering to data restrictions of the BL digitised books	A generalised methodology for preventing disclosure of restricted data from digitised books, applied to a short list of text analysis functions in DataSHIELD	University of Bristol, British Library, Content Mine, Researchers, text miners	
Phase 2: Proof of Concept	month 5 - month 7	Goal 2: Develop a proof of concept implementing findings of Goal 1 (remote analysis of text data using openly available digitised books from the British Library)	- Expand test infrastructure to include DataSHIELD methodologies for access-analysis of restricted data (in line with the findings of the advisory group) * Server side - install R parser and implement restrictions on the types of analysis that can be done. * Server side and client side - modify existing R functions (or create new functions) to conduct the analyses shortlisted in phase 1. >Test the functions using the open dataset from the British Library - the functions must adhere to disclosure, copyright and licensing restrictions of non-open digitised datasets held at the British Library. > Security test the infrastructure and functions - Test the new infrastructure using openly available British Library digitised book dataset.	Implement DataSHIELD methodology scoped in goal 1 for application to the digitised books. Build a proof of concept DataSHIELD text analysis package	Demonstrate remote restricted textual analysis using DataSHIELD methodologies.	University of Bristol, AMASED advisory group	
Phase 2: Proof of Concept	month 4 - month 7	Goal 3: Develop a proof of concept for the remote analysis of F1000 Research paper data.	- Adaptation of existing DataSHIELD infrastructure (based on a model scoped in Phase 1) to analyse openly available data provided by F1000 Research - Replicate an F1000 Research research paper analysis using the test DataSHIELD infrastructure - liaise with F1000 Research to identify a plan for a pilot implementation	Build a proof of concept of DataSHIELD for use in a data publishing setting. Create implementation plan for the pilot.	Demonstrate that remote analysis of openly available research paper data can be replicated using DataSHIELD	University of Bristol, F1000 Research	
Phase 2: Market Research	month 4 - month 7	Goal 4: Scope the web-based user interface for the software	- Interaction design team to liaise with DataSHIELD developers, the invited group of digital humanities researchers, and existing dataSHIELD users to scope a user interface	Produce a project report outlining scoping findings and suggested user interface	Model for design and implementation of a user interface for DataSHIELD	Good Form and Spectacle, University of Bristol, Researcher Advisors, current users	
Phase 2: Scoping	month 4	Goal 5: Scope the challenges of integration of the data cleaning tool ICT-RD (dataspring proposal 3.3) with DataSHIELD (www.datashield.ac.uk).	- Liaise with researchers at London Metropolitan University to establish a method to integrate ICT-RD data cleaning tool with DataSHIELD. Meeting booked 12/08/15, London	Identify a workflow and methodology to implement the ICT-RD data cleaning tool in existing DataSHIELD infrastructure for numerical data.	A defined and realistic methodology for integrating the data cleaning tool within DataSHIELD	University of Bristol, London Metropolitan University	
Phase 3: Proof of Concept	month 8 - month 13	Goal 1: Implement ICT-RD data cleaning tool (proposal 3.3) within a standard DataSHIELD infrastructure.	- Develop a DataSHIELD compatible package (or series of DataSHIELD functions) using the ICT-RD algorithms for data cleaning.	A prototype data cleaning package (or suite of functions) within DataSHIELD.	A data cleaning package integrated into core DataSHIELD software	London Metropolitan University, University of Bristol	
Phase 3: Evaluation and Implementation	month 8 - month 13	Goal 1: Test and evaluate the integrated data cleaning tool within DataSHIELD	- Define a group of data providers and researchers to test the data cleaning tools using a number of datasets including DataSHIELD simulated data and open paper data from F1000 Research - Feed the testing evaluation into continued development/maintenance of the data cleaning tool	Produce an automated data cleaning tool in DataSHIELD that will give the researcher and the data provider the quality status of their data.	Test users (researchers and data providers) use DataSHIELD to error and quality check their data. The tool flags data issues to the data provider.	London Metropolitan University, University of Bristol, data providers, evaluation team	
Phase 3: Evaluation and Implementation	month 8 - month 13	Goal 2: User evaluation of DataSHIELD on openly available text data (digitised books) from the British Library	- Invite a closed group of researchers to train on the system and beta test it for functionality using the openly available digitised books dataset from the British Library. - Evaluate the beta test and address feedback appropriately within DataSHIELD.	Evaluation of DataSHIELD by researchers	Feedback of user evaluation into the final infrastructure.	University of Bristol, British Library, group of digital humanities researchers,	
Phase 3: Evaluation and Implementation	month 8 - month 13	Goal 3: Implement DataSHIELD as an access-analysis method for non-open digitised books from the British Library	- Deploy DataSHIELD on a subset of non-open digitised books: * do this locally at Bristol with an access agreement in place from British Library to use the data * security test the infrastructure and functions - Implement the infrastructure developed by Bristol at the British Library * Security test the access-analysis and the stability of the system - Invite a closed group of researchers to train on the system and beta test it * Evaluate the beta test	Implementation of DataSHIELD infrastructure at the British Library	Analysis of non-open digitised datasets held by the British Library	British Library, University of Bristol, group of digital humanities researchers, evaluation team	
Phase 3: Evaluation and Implementation	month 8 - month 13	Goal 4: Implement DataSHIELD as an access-analysis method for sensitive paper data at F1000 Research	- Deploy DataSHIELD on a subset of sensitive F1000 Research paper data: * may need to do this locally at Bristol with an access agreement in place from F1000 Research to use the data? * security test the infrastructure and functions	Evaluation of DataSHIELD by data publishing stakeholders	Feedback of user evaluation into the final infrastructure.	University of Bristol, F1000 Research, test user group, evaluation team	

Phase	Months	Goal	Objectives	Deliverables	Indicators of Success	Work Assigned	Costing
			<ul style="list-style-type: none"> - Implement DataSHIELDwith F1000 Research for sensitive paper data * Security test the access-analysis and the stability of the system - Invite a closed group of stakeholders to train on the system and beta test it - Evaluate the beta test 	Implementation of DataSHIELD infrastructure at F1000 Research to analyse sensitive paper data.	Analysis of sensitive data within F1000 Research		