

ResearchSpace Business Requirements & Specification

Date: May 2010

Version: 2



Contents

1	Introduction	7
1.1	Document Purpose	7
1.2	Executive Summary & Walkthrough	7
2	ResearchSpace Processes – High Level	9
2.1	ResearchSpace System	9
2.2	Terminology	11
3	Concepts	13
3.1	Objectives	13
3.2	Business Rules	14
3.3	Main Technical Standards – Summary	15
4	ResearchSpace System Elements - Summary	16
4.1	Collaborative Content Management System	16
4.2	Social Networking Tools	16
4.3	Document / Asset Management	17
4.4	ResearchSpace and Images and other Digital Assets	17
4.5	Research Tools	17
4.6	Collaborative editing tools	17
4.7	RDF Stores	17
4.8	CMS Stores	17
4.9	Mechanism for RDF and CMS data synchronization	18
5	ResearchSpace Access and Roles	19
5.1	General Provisions	19
5.2	ResearchSpace Administrator	19
5.3	ResearchSpace Support	20
5.4	Project Administrator	20
5.5	Password Requirements	20
5.6	Audit	21
6	ResearchSpace Project Setup	22
6.1	Create Project Processes	24
6.2	Scope	24
6.3	ResearchSpace Governance & Administration	24
6.4	Project Request	24

6.5	Initial Access.....	25
6.6	Training.....	26
7	Environment Components.....	27
7.1	ResearchSpace Linked Data.....	27
7.2	Choosing what to publish?.....	27
7.3	Research Tools – General.....	28
7.4	ResearchSpace and Content Management.....	28
7.5	ResearchSpace and Social Networking.....	29
7.6	ResearchSpace Text Searching.....	29
7.7	ResearchSpace Auditing and Reporting.....	29
7.8	ResearchSpace and Project Web Sites.....	29
7.9	ResearchSpace Levels of Access.....	30
7.10	ResearchSpace and Co-reference.....	30
8	ResearchSpace Environment.....	31
8.1	Environment Processes.....	33
8.2	Introduction.....	33
8.3	ResearchSpace Administrator’s Dashboard.....	33
8.4	Project Administrator’s Dashboard.....	34
8.5	Project Team Member Dashboard.....	35
8.6	Workflow and Content changes.....	35
8.7	Workflow and Data.....	36
8.8	ResearchSpace Workflow – Configuration.....	36
8.9	Navigation.....	37
8.10	Collaboration Environments.....	37
8.11	Researcher’s Data Basket.....	38
9	ResearchSpace Research Tools.....	39
9.1	Process for using Collaboration and Research Tools.....	41
9.2	Summary.....	41
10	Research Tools – Detail.....	43
10.1	Introduction.....	43
10.2	Semantic Search Tool.....	43
10.3	Terminology Mapper.....	46
10.4	Data Annotation Tool.....	48
10.5	Image Annotation Tool.....	49
10.6	Image Zoom (and annotation).....	50
10.7	Image Compare.....	50
10.8	Relationship / Link Editor.....	50
10.9	Version Comparison (Track Changes).....	51
10.10	Geographical and Timeline Mapping.....	51

11	ResearchSpace Publication	53
11.1	Templates and Plug-ins	53
11.2	ResearchSpace Tools as public web site plug-ins	53
11.3	Security	53
12	ResearchSpace Data Gateway	54
12.1	Introduction	56
12.2	Conversion Templates	56
12.3	Online Data Entry	58
12.4	Asset Upload	58
12.5	Terminology	58
	Annex 1 – Research Tools Functionality	60
12.6	Image Annotation / Zoom Image Annotation	60
12.7	Data Edit & Data Annotation	62
12.8	Controlled Data Entry	64
12.9	Image Overlay	65
12.10	Terminology Mapper	67
12.11	Timeline Mapper	69
12.12	Geographical Mapper	71
12.13	Data Basket	73
	Annex 2 - Technical Architecture.....	74
13	Introduction.....	74
13.1	Architectural overview	74
13.2	General principles	75
13.3	ResearchSpace Components.....	75
14	ResearchSpace Authentication and Authorisation	76
14.1	Introduction	76
14.2	Purpose	76
14.3	Scope.....	76
14.4	Architectural Process.....	77
14.5	Interfaces	77
14.6	Additional Controls	77
14.7	Security	77
14.8	Question and Answer	78
15	Shared and Specific Stores.....	79

15.1	Introduction	79
15.2	Purpose	79
15.3	Scope.....	79
15.4	Architectural Process.....	79
15.5	Interfaces	80
15.6	Additional Controls	80
15.7	Security	80
16	ResearchSpace Registry.....	81
16.1	Introduction	81
16.2	Purpose	81
16.3	Scope.....	81
16.4	Architectural Process.....	81
16.5	Interfaces	81
16.6	Additional Controls	81
16.7	Security	81
17	ResearchSpace Catalogue.....	82
17.1	Introduction	82
17.2	Purpose	82
17.3	Scope.....	82
17.4	Architectural Process.....	82
18	Document and Digital Assets Store.....	84
18.1	Introduction	84
18.2	Purpose	84
18.3	Architectural Process.....	84
18.4	Interfaces	84
18.5	Additional Controls	84
18.6	Question and Answer	84
19	ResearchSpace Toolbar	85
19.1	Introduction	85
19.2	Interfaces	85
20	Co-reference Services.....	86
20.1	Introduction	86
20.2	Component Location	86
20.3	Interactions within the ResearchSpace environment	86
20.4	Inputs / triggers.....	86
20.5	Outputs / actions	86

21	Social Networking Tools – Main user area	87
21.1	Introduction	87
21.2	Component Location	87
21.3	Interfaces	87
22	Research Tools – Tools to operate over data and assets.....	88
22.1	Semantic Search and Explore	88
22.2	Component Location	88
22.3	Interfaces	88
22.4	Inputs / triggers.....	88
22.5	Outputs / actions	89
23	Semantic Update.....	89
23.1	Component Location	89
24	Relationship Mapper.....	90
24.1	Component Location	90
24.2	Interfaces	90
24.3	Inputs / triggers.....	91
24.4	Outputs / actions	91
25	ResearchSpace Text Searching Component.....	92
25.1	Introduction	92
25.2	Interfaces / Resources Affected.....	92
25.3	Implementation Options	92
25.4	Purpose	93
25.5	Number of Indexes.....	93
25.6	Usage	93
25.7	Index Update.....	93
25.8	Access	93
	Annex 3 – Component List	94
	Annex 4 - Reusable Open Source	104

1 Introduction

1.1 Document Purpose

1.1.1 This document is the initial specification for ResearchSpace (RS). It gathers into one place the core ResearchSpace business requirements, functional and technical specification necessary for the development of the ResearchSpace system. This document will be developed further as a result of issues and feedback during stage 3 of the project during which the first working environment will be built.

1.2 Executive Summary & Walkthrough

1.2.1 ResearchSpace will be a non-commercial and non-profit online environment in which scholars can carry out effective collaborative research. It will allow rapid creation of project environments with the following capabilities:

- The ability to configure an environment to address specific research or collaboration objectives.
- Harmonise data with other data from other organisations.
- Analyse the available data using a consistent set of collaboration and research tools.
- Share digital assets including documents, images, video and audio.
- Publish the results of the research on the internet using semantic linked data and digital web applications.

1.2.2 Instead of spending time and capital budget on establishing individual technology infrastructures for every new project, scholars can request online project space from ResearchSpace. Joining the ResearchSpace environment and community immediately provides access to a storehouse of data and assets provided by different cultural heritage and art organisations. In addition it will provide a set of integrated collaboration and analytical tools to operate on that data.

1.2.3 ResearchSpace harmonises data provided from different sources such that it behaves like one single database by making it conform to a high level ontology and by providing tools for mapping vocabularies. It is designed to help scholars discover the relationships and connections between the information held by different institutions, and as a result discover new information to advance scholarly knowledge. ResearchSpace will support an expanding set of data and metadata related to collection object records, archaeological sites, buildings and monuments, conservation and scientific data, and so on. This support will include tools to simplify the process of gathering and organising data from organisations, both small and large, regardless of geographical location.

1.2.4 The results and new data generated during research projects will become part of an ever-expanding ResearchSpace knowledge base for use by other projects and organisations. ResearchSpace is founded with the understanding of the need for organisations to share data as well as retain management and control over new work until it is ready to be released to the scholarly and wider community.

1.2.5 A project would therefore request a private area on the ResearchSpace environment. On successful registration it would configure a suitable collaboration environment and determine the parameters for research including the necessary tools, workflows and objectives. The project coordinator would then establish user accounts for other research team members and initiate the work to gather data from existing data already on ResearchSpace and upload additional data required for the project. The

ResearchSpace project intends to provide tools necessary to upload and convert data into the Semantic data schema used by the system, although some consultancy may be required by some projects to fully achieve this.

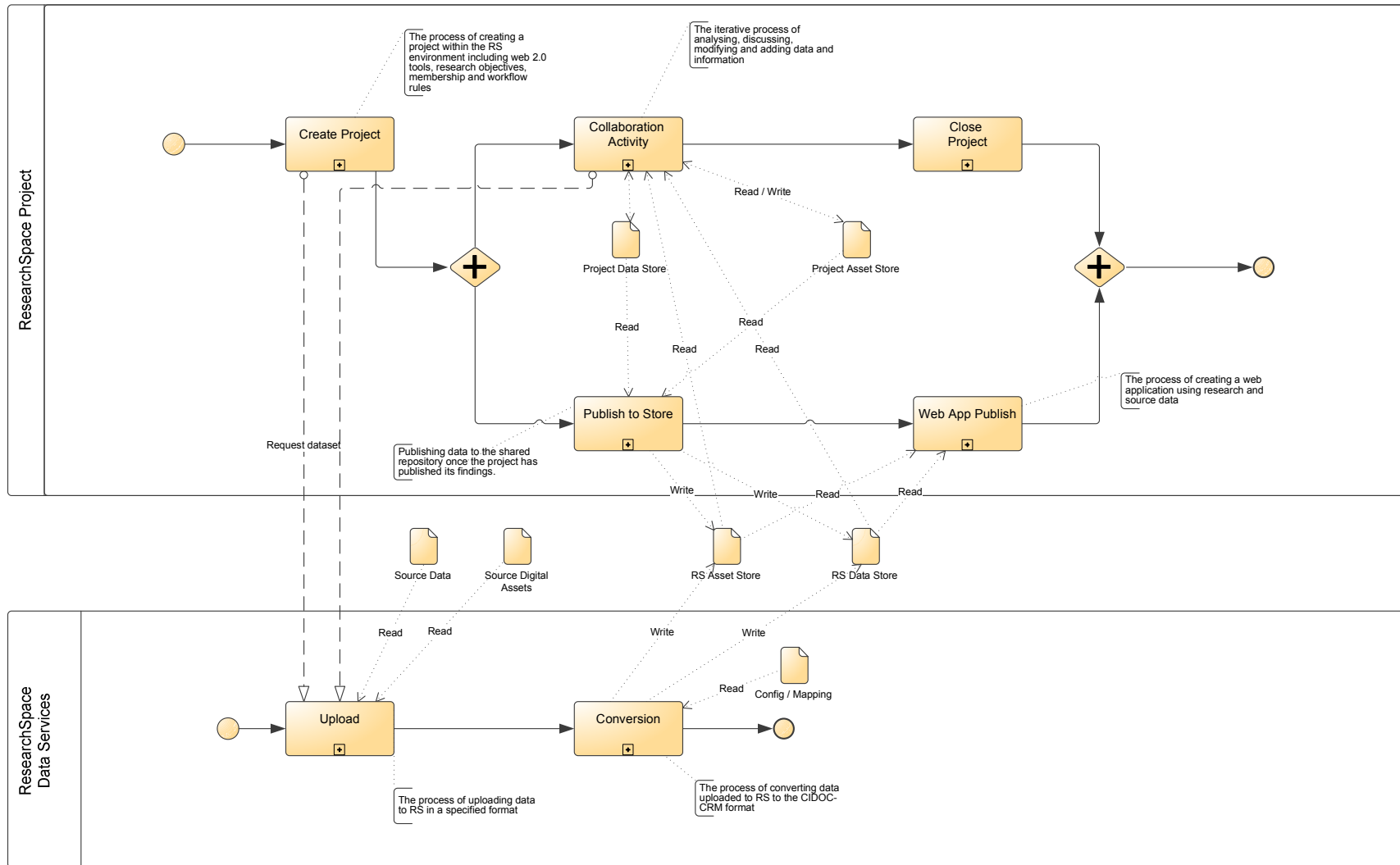
- 1.2.6 ResearchSpace tools will bring together functions commonly used by researchers including rich search and browse, data and multimedia annotation, time and geographical mapping etc. However, the tools will be in one environment and interface with each other using a common navigation system. This means that researchers would be able to switch between different tools without interrupting the focus and avenues of research. Other researchers can follow these paths and use the same tools to contribute to a different investigation in the same or different project. New tools and editors could also be developed by projects which could then be released for use by the larger ResearchSpace community.
- 1.2.7 The ResearchSpace environment will encourage knowledge generation by creating a seamless integration between data, research tools used to analyse the data, and collaboration applications. In this type of integrated environment research objectives can be discussed with immediate and direct access to information resources. In addition, project principals and project members will be provided with dashboards aimed at providing the latest information about project activity so that project momentum can be maintained over the project's lifecycle. The dashboards will be developed specifically to help projects monitor and manage their activity within ResearchSpace.
- 1.2.8 ResearchSpace will help projects keep track of the information by maintaining a catalogue of all data and assets. The ResearchSpace catalogue will also record the status of information so that decisions can be made throughout the project that will allow rapid assimilation of project outputs. This in turn will promote more effective and accurate publication of data. The use of RDF data will be comprehensive and will be used for supporting research, publication and presentation, including web applications. The use of RDF will serve the purpose of minimising barriers for bringing cultural heritage information together, supporting more sustainable research, and reducing the costs involved in collaborative research. The tools provided by ResearchSpace will therefore provide functionality that makes full use of the potential of semantic technology.

2 ResearchSpace Processes – High Level

- Create Project – The Process of establishing a project space on ResearchSpace
- Collaboration Activity – Use of the tools within the ResearchSpace collaborative environment
- Publish to Store – The publication of data from the project store to the ResearchSpace store
- Web Application Publication – The generation of a web site to publish project outputs
- Upload – The process of transferring data and information to the ResearchSpace environment, either to ResearchSpace shared repositories or specific private repositories.
- Conversion –The process of converting data to the CIDOC-CRM semantic format
- Close Project –The process of decommissioning a particular project space.

2.1 ResearchSpace System

- 2.1.1 ResearchSpace will be constructed using two (or more) off the shelf open source systems with three main elements. These elements are a RDF database management system, a Content Management System and a Digital Asset Management system. The system will support document and multimedia management as well as the tools necessary to support collaboration. ResearchSpace will be developed to provide a number of research tools designed to work with RDF data. These tools will operate seamlessly with social networking tools purposed to support collaborative research.
- 2.1.2 ResearchSpace should provide administrators and project coordinators with a simple and, if possible, unified mechanism for assigning access rights within projects. For this, ResearchSpace project coordinators will be provided with dashboards providing information about the project, its contributors, activity and communication between different project members.



ResearchSpace System		author: Dominic	created: 05/12/2010 21:01:26	THE BRITISH MUSEUM
High level data flows		version: 1.0	modified: 27/05/2011 10:33:38	
		status: created		
rsflow (3).vsd				

2.2 Terminology

- 2.2.1 **ResearchSpace Data Store:** - The store of data in 'RDF' format available to all projects and which is also available as 'linked data' for the community.
- 2.2.2 **ResearchSpace Document Store:** - The store of documents (PDF, MS Word files, etc) available to all projects that use the environment. Documents can also be published to the community through project web applications.
- 2.2.3 **ResearchSpace Asset Store:** - The store of digital assets (images, video, audio etc) available to all projects that use the environment and the internet community through linked data interfaces and web applications.
- 2.2.4 **Project Data Store:** - The store of data in 'RDF' format available only to a specific ResearchSpace project.
- 2.2.5 **Project Document Store:** - The store of documents (PDF, MS Word files, etc) only available to a specific ResearchSpace project.
- 2.2.6 **Project Asset Store:** - The store of digital assets (images, video, audio etc) available only to a specific ResearchSpace project.
- 2.2.7 **Project Data Environment:** - The store of digital assets (images, video, audio etc) available to a specific ResearchSpace project.
- 2.2.8 **Data:** - The structured metadata and data related to cultural heritage objects, buildings, sites etc.
- 2.2.9 **Digital Assets:** - The digital assets that are uploaded to ResearchSpace digital asset stores.
- 2.2.10 **Digital Documents:** - The electronic documents that are uploaded to ResearchSpace document stores.
- 2.2.11 **ResearchSpace Registry:** - A record of information about ResearchSpace projects, including the participating people and organisations.
- 2.2.12 **ResearchSpace Catalogue:** - A record of data and assets contained within ResearchSpace together with the location and status of that data.
- 2.2.13 **Co-Referencing Services:** - Services that can map the equivalence of data references as well as manage their deprecation.
- 2.2.14 **ResearchSpace Dashboard:** - The Dashboard available to ResearchSpace administrators which provides information on the activity of projects and tools for its administration and organisation.
- 2.2.15 **Project Lead Dashboard:** - The Dashboard available to the people leading a project providing information on the activity of the specific project and tools for its administration and organisation.
- 2.2.16 **Project Member Dashboard:** - The Dashboard available to a project research team which provides information on the activity of the project, personal activity and tools for help with the use of the environment.
- 2.2.17 **ResearchSpace Templates:** – The data templates that reflect different metadata models that have been mapped to the CIDOC-CRM structure.

- 2.2.18 **Online Data Entry Screen:** – The data entry screens that reflect the ResearchSpace data templates which are mapped to CIDOC-CRM.
- 2.2.19 **ResearchSpace Gateway:** – The infrastructure that consumes and ingests data into ResearchSpace which includes web services and online data entry forms.
- 2.2.20 **RS Unstructured File:** – A document or file with irregular non-computer readable data.
- 2.2.21 **RS Structured File:** – A file containing consistent computer readable data.
- 2.2.22 **RS Template Compliant Data:** – A structured data file or data source that complies with a ResearchSpace template which has been mapped to CIDOC-CRM.
- 2.2.23 **ResearchSpace Linked Data Service:** – The interface that allows data from the ResearchSpace shared repository to be served as linked data to the Internet.
- 2.2.24 **ResearchSpace Data Basket:** – A mechanism for collection data and links for use in research and collaboration tools.

3 Concepts

3.1 Objectives

- 3.1.1 To create an RDF infrastructure based on the CIDOC-CRM ontology that harmonises data contributed by different cultural heritage and arts organisations. ResearchSpace will initially bring data from different sources into one central system as a resource for online collaborative research. A future objective, dependent upon the development of semantic technology and its adoption, would be to be able to use ResearchSpace tools over decentralised RDF stores.
- 3.1.2 To provide collaboration environments and tools to enable effective online scholarly communication with direct access to supporting data. The main issue with current online scholarly collaboration is that the data is not, 'on hand'. The ability to draw upon source materials during research is a major enhancement to this type of environment.
- 3.1.3 To provide research tools that harness the benefits of semantic data for the discovery of new information and knowledge and support a range of research methods and techniques for analysing data, information and multimedia resources. With the introduction of semantic frameworks the connection between data and artefacts from different sources can be established allowing scholars to draw upon them without laborious manual investigation.
- 3.1.4 Allow publication of data and documentation in web applications, and as linked data, which would be available to the scholarly community working outside the ResearchSpace environment.
- 3.1.5 Be accessible to a wide range of humanities organisations, large or small and regardless of internal resources.
- 3.1.6 Provide a full web browser interface, and open technology architecture to encourage community development.

3.2 Business Rules

- 3.2.1 **Business Rule 1: Original Research** - Data and assets generated by research are private to the project and restricted to project members until published to the shared ResearchSpace repositories. Projects would be expected to publish this data to the shared repository at some stage (as soon as practicable and reasonable) and data would be shared on the repository at the end (decommissioning) of the project.
- 3.2.2 **Business Rule 2: Established Data** – Existing data and assets, for example accessible organisational collection data and associated images, should be uploaded to the shared ResearchSpace repository and be available to all projects. Some multimedia assets will require some access restrictions and this reflects the practical reality that some media assets are more likely to be subject to conditions and restrictions. However, ResearchSpace would always encourage free and unfettered access to cultural heritage data.
- 3.2.3 **Business Rule 3: Open System Standards** - the ResearchSpace tools should not be dependent on proprietary APIs. It should be possible to take any tool developed for use with open standard RDF data and incorporate that tool into ResearchSpace without substantial redevelopment. Similarly, tools developed specifically for ResearchSpace should be useable with little or no modification outside the ResearchSpace environment. This business rule ensures the open nature of ResearchSpace and keeps the RDF research tool model simple and accessible.
- 3.2.4 **Business Rule 4: Image Restrictions** - Images will require some security so that specified images can be hidden from some users but be available to others. This reflects the practical reality that some images are more likely to be subject to conditions and restrictions. It is anticipated that standard Digital Asset Management tools can be used to control image access without offending other ResearchSpace business rules. ResearchSpace will encourage sharing of digital media whenever possible.
- 3.2.5 **Business Rule 5: Ontology Standard** - In order to obtain a good level of data harmonisation and therefore allow effective exploration of data supplied from different sources, the ontology standard for all data will be the CIDOC Conceptual Reference Model (CRM).
- 3.2.6 **Business Rule 6: Use Cases** – The main ResearchSpace elements of data, collaboration, analysis, and web publication should be both integrated but also available as separate functions to encourage a wide range of project and related use.
- 3.2.7 **Business Rule 7: Open Source** – New software created will be released to the community freely as open source. Existing open source tools should, if possible, be utilised for ResearchSpace development.

3.3 Main Technical Standards – Summary

3.3.1 The main technical standards for the project are:

- CIDOC CRM - Conceptual Reference Model
- ICOM – International recommendations on linked data
- RDF – W3C open data standard, Resource Description Framework
- SOA – Service Orientated Architecture
- CMIS – Content Management Interface System
- MPEG-7 – Multimedia metadata standard

annotation with the appropriate associations or relationships in the RDF store. By default, CMS tools only store data in the CMS database and therefore some work would be required to support this alternative commitment of data.

4.3 Document / Asset Management

4.3.1 It is expected that document and digital asset management functions are provided by the chosen content management system. However, these assets will require description and themselves be annotated as part of research analysis and activity. Asset metadata and annotations should therefore also be part of the RDF store. In order to maintain independence from CMS API dependencies the selection criteria for the ResearchSpace CMS should include the ability to search the asset repositories in a way not reliant on proprietary CMS code.

4.4 ResearchSpace and Images and other Digital Assets

4.4.1 ResearchSpace also stores images and videos. Many of these images will be of objects as described in the object records. ResearchSpace would need to manage high resolution images and allow ResearchSpace users to zoom into the detail of a represented object, painting or other images. Research tools should be created to allow users to find, edit and annotate images at different levels of zoom, i.e. embed annotations at different levels of magnification. Although a Digital Asset Management system will be used to manage images, image metadata must also be stored and linked to other metadata within the RDF stores.

4.5 Research Tools

4.5.1 This specification describes the main research tools that should be available in the first production version of ResearchSpace (although some detail may be part of an additional document). However, the architecture of ResearchSpace should be such that projects can easily develop other research tools that can also be used by other ResearchSpace projects. These tools will read and write to the RDF store.

4.6 Collaborative editing tools

4.6.1 This will depend upon the tool. For example, for shared document creation/editing users would be able to create a document/text for storage in a document library.

4.7 RDF Stores

4.7.1 The RDF stores are all named graphs (distinct datasets with separate security) within the same RDF database. The search system for researchers must therefore be able to search across two or more different RDF named graphs including the user's current project graph, the RS shared graph and other administrative stores. When data is asserted into these stores this is recorded in another named graph called the, 'RS Catalogue'. The 'RS Registry' which records information about the different projects is another named graph.

4.8 CMS Stores

4.8.1 The chosen content management system will have facilities for asset storage including both documents and digital assets. The RS Catalogue will also record items that are placed in, and removed from, these asset stores using appropriate metadata. This means that users should be prompted to enter metadata about these assets. The mechanism for this may be the CMS system itself, in which case this data should be replicated in the RS Catalogue. The metadata should include the user

(derived from the login and associated user profile), the document title, version, and brief description and, if applicable, the identifier of the data record that it represents. The upload interface will provide the opportunity to include this information.

4.9 Mechanism for RDF and CMS data synchronization

- 4.9.1 The development of ResearchSpace should avoid customisation of third party code wherever possible and prefer mechanisms that could easily be transferred to another open source information system. As such the preferred method of transferring data between the CMS store (likely to be a relational database) and a RDF store is either through an available SOA API facility or at the database level, rather than altering third party code. It should be possible to upgrade third party components used within ResearchSpace with little or no conflict with bespoke elements of the system.

5 ResearchSpace Access and Roles

5.1 General Provisions

5.1.1 ResearchSpace user account access should be role-based but needs to provide authentication and authorisation for different systems, the collaboration (CMS) environment and the RDF database environment. When users logon to the ResearchSpace environment it is likely that they will use the CMS authentication system. The system requires that on authentication to the CMS that this in turn authenticates against the RDF database system. This may be achieved using the following mechanisms:

1. The usernames and passwords for both systems are synchronised. When a username and password is setup it is created in both authentication systems. When a password is changed it needs to be changed in both systems.
2. That both systems subscribe to a single authentication system, such as a directory service. This will depend on whether both systems are, or can be made, LDAP compliant which is the preferred method for authentication, authorisation and profile storage.
3. It should be possible to create other roles with different access rights. ResearchSpace linked data will be available to all Internet users.

5.1.2 The following user rights are required (as well as the ability to create others)

Role	Project Forums	Project Stores	ResearchSpace Stores	Other
ResearchSpace Administrator	Administrator	Administrator	Administrator	Access to project forums and stores will be determined by acceptable user policies. Has access to ResearchSpace dashboard and tools.
ResearchSpace Support	Administrator	Administrator	Administrator	Support for ResearchSpace tools and data issues
Project Administrator) (Lead)	Administrator	Administrator	Write	The user is a full participant in the project and the data analysis and generation. The user has access to the Project Lead dashboard and associated functionality
Full Project Team Member	Write	Write	Write	The user is a full participant in the project and the data analysis and generation. The user has access to a personal dashboard.
Collaboration Team Member	Write	None	None	The user has no access to a data store and can only use the collaboration tools, but not the research tools.
Project Observer	Read	None	None	The user is a guest to the discussion forum and there is no data store for the project or the user has no access.

5.2 ResearchSpace Administrator

5.2.1 An account associated with the management of a particular site running ResearchSpace. The account provides facilities to manage the activities of projects that require central agreement such as:

- Project setup, allocating a Project Administrator account, and setting up the basic infrastructure;
- Project close down;
- Allocation of resources to a project.
- It also maintains oversight of the central assets.
- The system must provide the ability to carry out management decisions, and provide the appropriate information to ensure the account holder can carry out the role.

5.3 ResearchSpace Support

5.3.1 The support account provides for the administration of the tools and components of ResearchSpace. For example, it will be able to:

- Install ResearchSpace components;
- Configure all aspects of ResearchSpace;
- Install new research tools as requested;
- Install new collaborative tools as requested;
- Run the utilities to convert source data to RDF;
- Write any RDF into any ResearchSpace repository.

5.4 Project Administrator

5.4.1 A security role with access to all features of the system. This includes creation of the initial project environment and user administration. The project environment should still be straight-forward to create and provide a wizard type interface so the Project administrators can make all decisions necessary to create the project environment. These steps would involve:

- Establishing access policy to the site
- Set up of user and user notification
- Choosing Social Networking components
- Making workflow decisions
- Setup and configuration of the Document management library setup
- Configuration of Asset Management
- Creating the research questions and objectives
- The ability to move (publish) assets from Project stores to ResearchSpace stores

5.4.2 **Project Team Member** – an account that allows the user to contribute to the collaborative and research forums and enter data into the RDF store.

5.3.4. **Project Team Collaborator** – an account that allows the user to contribute to the collaborative tools (discussions) but not to the research tools.

5.4.3 **Project Observer** – an account that allows the user to read project information and collaborative discussions but without the ability to contribute to discussions.

5.5 Password Requirements

5.5.1 Passwords should be encrypted and use at least 8 characters containing lower and uppercase letters and include a number. If a user submits a password incorrectly three times in a row the user will be locked out and require a reset by the project administrator(s). Project administrators(s) should have the ability to unlock accounts and reset passwords. A lockout should be audited the project lead then

notified on the Project Lead Dashboard.

5.5.2 The username should be the email address of the user. This should be associated with their full name and contact details, and institutional affiliation. The user should be able to enter a profile screen available to that user and edit profile information. The following also form part of the specification:

- The system should use appropriate single sign-on authentication.
- The RDF subsystem will control access using simple "htaccess", informed by the authenticating system.
- Research tools should conform to the authentication system and be LDAP compliant but also be able to operate in a standalone environment.

5.6 Audit

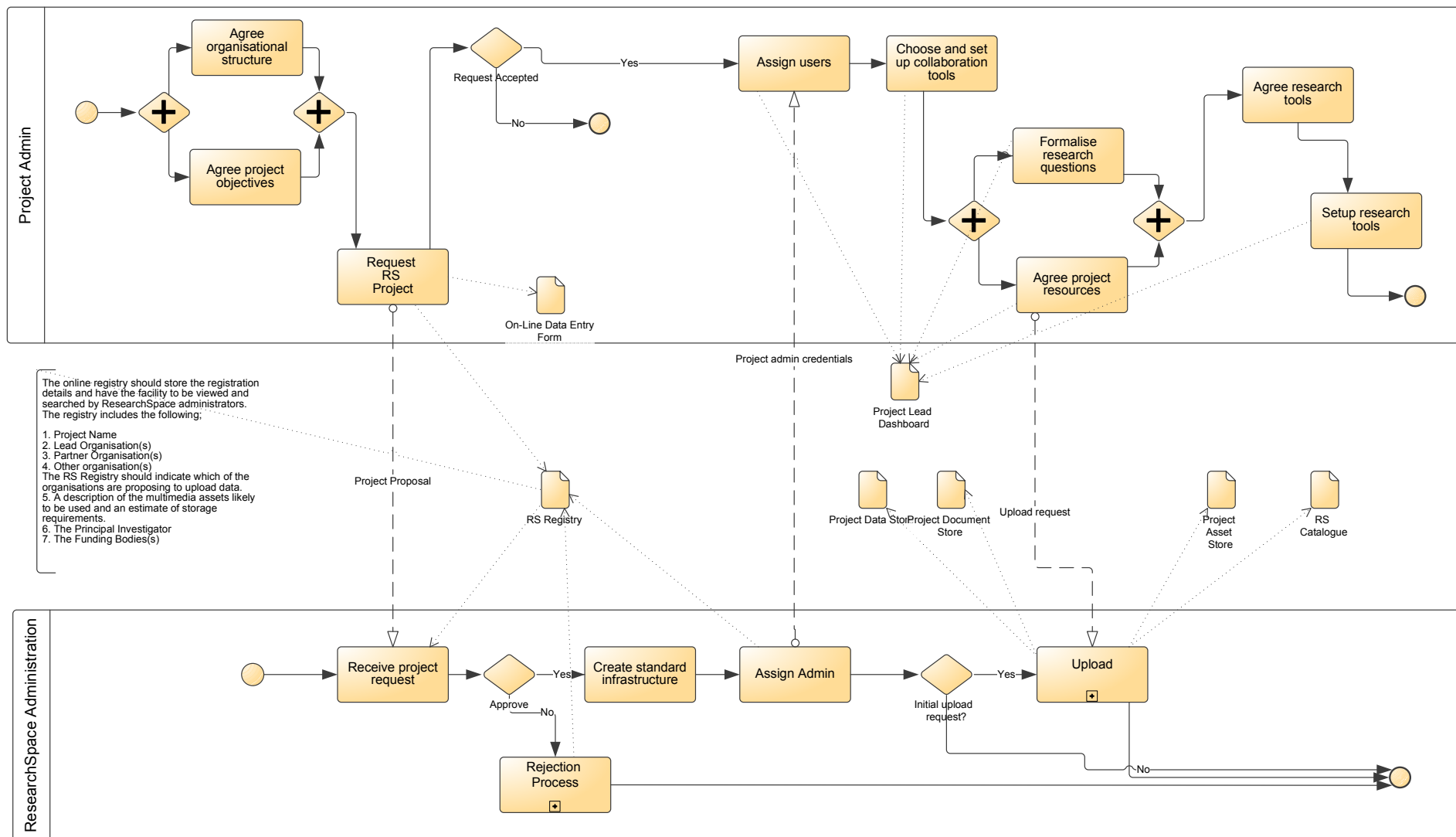
5.6.1 Collaboration posts (social networking contributions) should be recorded by the CMS system under the users account with appropriate date and time stamps. If data from the repositories is brought into a collaboration area then the CMS should also record the source of the information.

5.6.2 Data entry and annotations to the data and assets should be recorded in the RS Catalogue using the appropriate account details and date/ time stamps.

5.6.3 Upload of documents or digital assets.

5.6.4 The system should record the operations with data as follows:
User and date details for documents and assets (1) uploaded, (2) created, (3) modified and (4) deleted in the asset stores.

6 ResearchSpace Project Setup



2 Creating a ResearchSpace Project

ResearchSpace System		author: Dominic	created: 05/12/2010 21:01:26	THE BRITISH MUSEUM
Create Project		version: 1.0	modified: 27/05/2011 10:33:38	
		status: created		
		rsflow (3).vsd		

6.1 Create Project Processes

- Agree organisational structure – the project organiser, usually the principal investigator, brings together all the elements needed to support the project.
- Agree project objectives – required to give focus to the collaboration.
- Request RS project – request, through online form, project space on ResearchSpace.
- Receive project request – project request received by the ResearchSpace administration.
- Rejection – a reject request to ResearchSpace.
- Create standard infrastructure – creation of the basic project space for the requested project.
- Assign administration – transfer of project administration to the requested project.
- Assign users – Create user accounts for the project team members.
- Choose and setup collaboration tools –configure the basic project space as required for the particular project.
- Formalise research questions - seed the collaboration areas.
- Agree project resources – including accounts and storage requirements.
- Setup research tools – select and configure the tools needed for the project.
- Upload – A separate process for uploading documents, assets and data.

6.2 Scope

- 6.2.1 This process captures the creation of a project within the ResearchSpace environment, which involves the collaboration between those proposing the project and the ResearchSpace platform administrators. It begins with a person or organisation proposing a project, and ends when the completion of the administration arrangements so that research proper can begin.

6.3 ResearchSpace Governance & Administration

ResearchSpace will require a low overhead framework for management and administration of the environment. The details of this framework will be developed by the ResearchSpace project through an advisory committee made up of representatives of organisation likely to make use of the environment.

6.4 Project Request

- 6.4.1 ResearchSpace will provide an online registration form for the purpose of requesting space and providing information about the project. The registration form will have the following fields for completion by the project personnel.

- Project Name – Each Project should have a unique RS identifier*
- Project URI – If the project has an Internet domain name
- Description of the Project and Project approach
- Lead Organisation(s) – There could be one or more lead organisations per project*
- Partner Organisation(s) – There could be one or more partner organisations actively involved
- Other organisation(s) – There could be other contributing organisations not actively involved in the project but contributing data or occasional input
- (Address details for all organisations)
- The Principal Investigator(s)*
- Principal Investigator(s) Email Address(es)*
- Principal Investigator(s) Telephone(s)*
- Project Contract name
- Contact Email Address

- Contact Telephone
- Business Address Details
- The Funding Bodies(s) – could be financed by partner-institutions
- One contact detail for each funding body
- Indicate which organisations are proposing to upload data.
- Estimate extent of multimedia assets
- Overall storage estimate
- Estimated Project Closure Date

(* = mandatory)

6.4.2 The request is saved to the RS Registry and appears in the RS Administrator’s dashboard. The request is also emailed to pre-configured email accounts. On approval the instance is created, and assigned to the Project Administrator, who is given the credentials to logon to ResearchSpace and manage the project. The RS Registry will be completed by ResearchSpace personnel recording the following;

- Decision of ResearchSpace Authority (Accept, Reject, Pending).
- Decision Notes.
- Project Status (ongoing).
- Status Notes.
- Project Subscription Amount.

6.4.3 The ResearchSpace Registry should be available for public viewing although it should be possible to configure which Registry fields are included.

6.5 Initial Access

6.5.1 The Project Administrator will sign on to ResearchSpace and access the dashboard system. The Project Administrator’s dashboard should have the facility to perform the following;

- **Provision Users** – The Project Administrator can set up accounts which should initiate an email to team members to register with the project. Each team member will be required to provide contact details which should be saved to the LDAP directory service. This should include organisational and alternative contact details.
- **Choose Collaboration Tools** - The Project Administrator would select from a list of collaboration tools and initiate the subject matter or questions that would determine the different threads of work to be undertaken by the team and the Project Administrator should be able to add additional questions or areas for research at any time. For example, if four research questions are set for a discussion forum then these should be created in the project space with links to them from the user’s dashboards. The same should occur in terms of subject areas for blog and wiki development. The Project Administrator should be able to add additional tools as and when needed. It is expected that the CMS would provide tools for alerts and general communication to the project researchers. For example, when a project objective or question is set, this would automatically trigger an email to the project team. ResearchSpace would also provide services for individual members of a project team including a personal task list.
- **Choose Research Tools** – The Project Administrator can select from a list of registered research tools. These tools will appear in the project research toolbar which should be present within the collaboration environments. By default the toolbar will include all the basic ResearchSpace tools described in this document. As new tools are developed Project Administrators can decide whether to use them.

6.5.2 During this period the Project Administrator will arrange for uploads of data and information for the project from the contributing organisations. The upload mechanisms are described later in this document.

6.6 Training

The ResearchSpace environment will be designed with non-technical users in mind. However, some training and guidance will be required for certain aspects of the environment. The development project will produce online manuals, training materials and video tutorials aimed at the different technical, administrative and user roles.

7 Environment Components

7.1 ResearchSpace Linked Data

7.1.1 The ResearchSpace project supports linked data and would provide the facility to publish collection data, according to established linked data protocols, to the internet (as well as a SPARQL EndPoint). ResearchSpace supports W3C Semantic Web Standards and the ICOM guidelines for URI provisioning. As such, contributors to the ResearchSpace shared data repository would have the option to publish their data as 'linked data' and make it available to others on the Internet. Source data would be published using the organisations own URI schema.

7.1.2 As individual research projects progress they would use, annotate and create new versions of the data provided by the contributing organizations. When the project has concluded its work and published its findings, the project should also make this data (and will be encouraged to do so) available as linked data using the schema of the data that it refers to. This means that when consulting a URI, for example, the 'Rosetta Stone', the body of work generated by ResearchSpace about the Rosetta Stone (including the original British Museum data) will also be available and clearly attributed. This is a potentially exciting consequence of using linked data and allows researchers to juxtapose their research findings alongside the source information making it more prominent and accessible to others but with clear attribution information.

7.1.3 It is also expected that ResearchSpace will host data for other (partner) organisations, in particular hosting collection and other cultural data as Linked Data. Sometimes this will be as part of ResearchSpace, but it may also be that the organisation would like ResearchSpace to be the primary publisher, which would appear within the organisation web site. The following are specified;

- ResearchSpace will provide an RDF subsystem that adheres to Linked Data best practice, with resolvable URIs for all resources over HTTP (How to Publish Linked Data on the Web - <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>).
- The URIs will conform to the emerging ICOM guidelines for URI provisioning.
- The initial part of URIs for ResearchSpace-published data will typically be [http://researchspace.org/\[project-name\]](http://researchspace.org/[project-name])
- URIs should not change as they move from project space to ResearchSpace.
- URIs for external organisation will be supported, for example <http://collection.britishmuseum.org/....>, <http://site.britishmuseum.org/....>, <http://building.britishmuseum.org/....>
- ResearchSpace will provide a SPARQL endpoint that supports the standard of the time, currently <http://www.w3.org/TR/rdf-sparql-query/>, although this is under imminent revision.
- Organisations involved in a project can download RDF files kept in the project document store.
- Change of data in the RDF Store triggers a change of data in the CMS where necessary.

7.2 Choosing what to publish?

7.2.1 What data generated as part of the project would be published as linked data? Typically a project would want to publish the results or outcomes of the research. However, in the course of a project different versions of data may be generated. This means that, when it comes time to publish, a decision would be required on the final version.

7.2.2 Decisions about what version of the data generated by a project should be published would be far easier to assess on an ongoing basis rather than at the end or closure of a project. However,

ResearchSpace should provide tools for both approaches. This means that the Project Administrator should be alerted when annotations and new versions of data are added and this would be through the dashboard facility. In addition, an appropriate status should be assigned to indicate the basis on which the data is proposed. The terminology for this should be configurable per project. The proposal of the data or information should also be able to provide some contextual notes which should also be available when viewing the proposal.

- 7.2.3 The provision of status flags and comments should also be available to a Project Administrator in a screen report format. The Project Administrator should be able to go through all the changes with the ability to apply a status change to the data which would indicate the data and information that is accepted and will be published (whether for a research paper or article or through ResearchSpace digital publishing.)
- 7.2.4 For example, the British Museum publishes its data to ResearchSpace. A ResearchSpace project is established to research a particular type of object and finds relevant British Museum object records. Subsequent research shows that the place of production for a particular object is, in the opinion of the project, incorrect. A user from the project inputs a new version of the production place data. The object now has the British Museum's original version and the project's new version. The British Museum may still wish to publish their original version within their own web application. The project may wish to publish using its own version of the production place. The two are not incompatible.
- 7.2.5 In terms of linked data the British Museum would use, [http://collection.britishmuseum.org/object/\[inventory ID\]](http://collection.britishmuseum.org/object/[inventory ID]). If projects generate different versions of the production place, then this would use a project URI but with an appropriate schema extension to denote the status of the version. It should also provide an attribution for the person who submitted it. The two versions would be linked by an, 'rdfs:seeAlso' or 'owl:sameAs'.

7.3 Research Tools – General

- 7.3.1 External users of the system should be able to create applications against the linked data and SPARQL EndPoint. However, tools used within the ResearchSpace environment would be developed overtime by different organizations and projects. Tools developed by projects should become available to all other projects within ResearchSpace. Other tools that are developed for CIDOC-CRM / RDF data independently of ResearchSpace should also be easily incorporated into the ResearchSpace system.

7.4 ResearchSpace and Content Management

- 7.4.1 Some ResearchSpace tools require interaction with the Content Management System. The following provides a categorization of the different tools and their dependencies on the CMS.
- Tools that relate to the analysis, modification and enhancement of the collection data. These should have no dependency on the CMS system. The exceptions are the CMS tools themselves where the data created within them may also be required for committal to the RDF store where appropriate customisation will be required.
 - Tools that relate to the publication of web sites that utilise the data stored in the RDF store. These tools can make use of CMS functionality to publish web sites and web pages containing RS data.
 - Orchestration of ResearchSpace tools can utilise CMS workflow and business process features. Research tool configuration should be built into the tool and be independent of the DAM.
 - Data ingest can use CMS business process features.

7.5 ResearchSpace and Social Networking

- 7.5.1 Social Networking tools provide the foundation for scholarly collaboration within ResearchSpace. Therefore they need to be functional and intuitive environments on which to base other research activities. A post or submission within a social networking tool would also represent a potential data annotation for the RDF store.
- 7.5.2 It should be possible to specify that a post, or part thereof, is also being saved as an annotation against a particular record or record element in the RDF store. When making and identifying some text in a post for annotation against other data in the RDF store a reference will be needed which may require a lookup. The alternative is that a researcher must find the relevant object data first, launch an annotation tool and then bring any data they have created back to the post for insertion within the particular social networking forum.

7.6 ResearchSpace Text Searching

- 7.6.1 ResearchSpace will employ a search engine to deal with searching through text contained in more descriptive fields and created as a result of annotations and social networking posts. Search engine results should be compatible with the RDF Research tools expecting RDF results.

7.7 ResearchSpace Auditing and Reporting

- 7.7.1 The ResearchSpace system should provide a system for reporting the activity within a particular project. Although annotations (and associated attribution) would be part of the RDF store, it may be convenient for some data to be written also to the CMS database system (for activity reporting purposes) so that project reports can be displayed and manipulated easily.
- 7.7.2 ResearchSpace should provide a level of reporting functionality which will, in particular, allow a project to review work in a structured way. This means that project reports should be able to separate out different types of data input and provides 'sort' and 'filtering' ability on people, subjects, dates, objects etc. ResearchSpace would provide the ability to provide a summary of all the activity with links back to the detail, showing collaboration activity and data activity both separately and together.
- 7.7.3 For example, half way through the project the Project Lead decides to view activity to date. A report is compiled showing, in date order, all the activity - posts in the social networking tools and commits of data to the RDF store. Both are represented in a summary form but with links to expose the data (perhaps launching the original tool used to submit it). The report provides the option of removing the social networking comments and viewing only the RDF database activity and vice versa. The interface would also allow a report to show the activity of particular project team members and team members from particular organisations and filter using other metadata.
- 7.7.4 There should also be reporting facilities to provide management statistics about the use of the environment as a whole. This includes aggregate project logs which can be used to drill down into further detail.

7.8 ResearchSpace and Project Web Sites

- 7.8.1 ResearchSpace will provide tools for creating a web site and publishing data in a human readable form. The system should provide a number of different templates that can be further styled and customised.

7.8.2 ResearchSpace should provide a number of flexible plug-in tools for common web data representations (for example, see the British Museum's Collection Online system). Additional plug-ins would be developed by individual projects to suit particular needs but it should be possible to provide a reasonable level of functionality to display metadata and images and provide good navigation and searching.

7.9 ResearchSpace Levels of Access

7.9.1 The aim of ResearchSpace is to allow different organizations and experts to effectively participate in a collaborative research project. In the case of organisations like the British Museum, metadata will be uploaded in its entirety for use in any ResearchSpace project. Other organisations will be encouraged to take this approach and take advantage of ResearchSpace linked data services, particularly if they are unable to publish independently.

7.9.2 Data security on images will be important to many organisations who wish to protect income revenue through the exploitation of high resolution images. It is also possible that some images may be sensitive just in terms of their content. It is therefore important that ResearchSpace is secure as far as possible and provides strong password encryption and access security. Initial access to the environment would be through the Content Management System which should incorporate a strong security model.

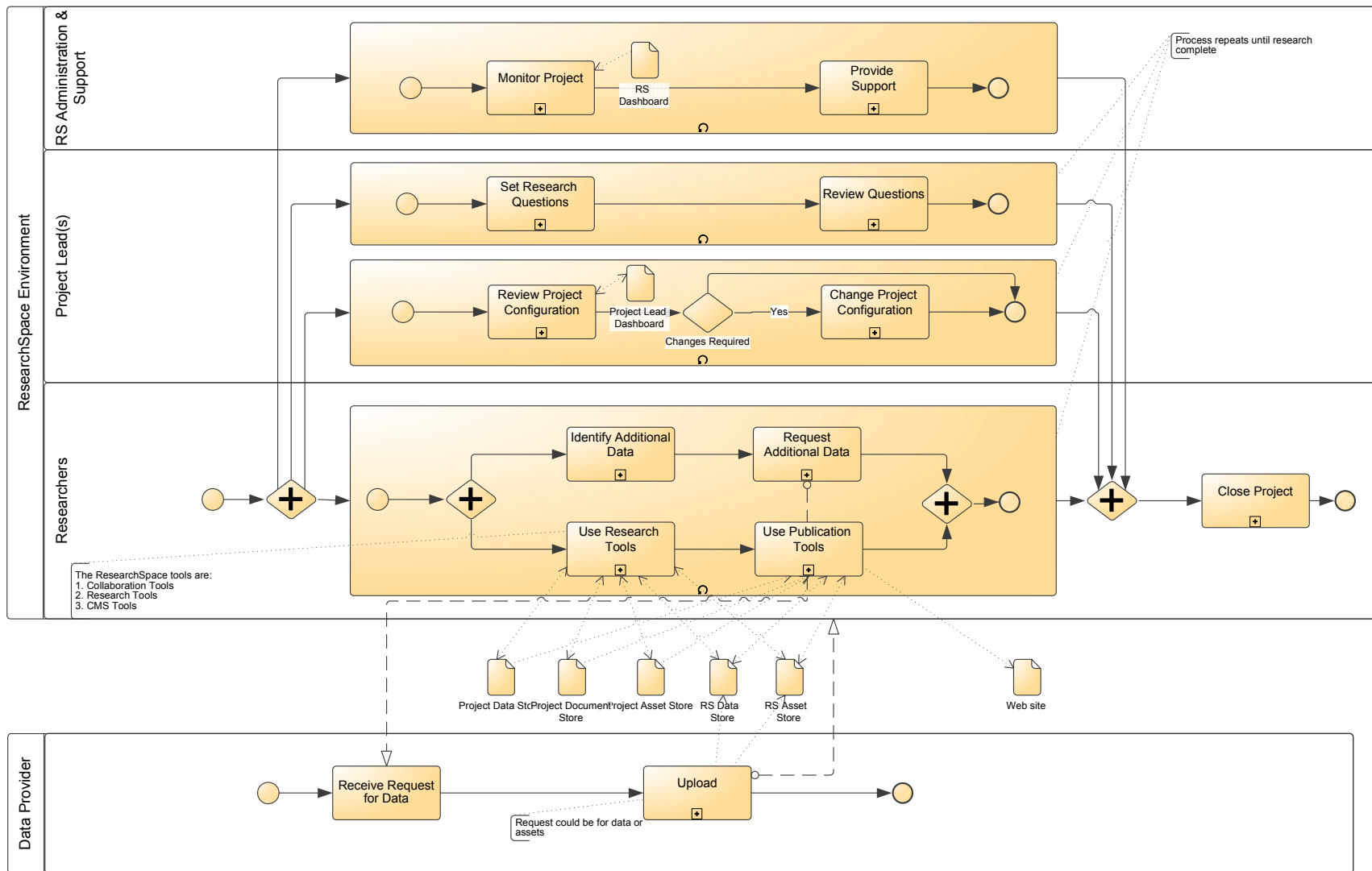
7.10 ResearchSpace and Co-reference

7.10.1 When different organizations upload data for a particular project, although the data will be organized according to the CIDOC-CRM, the taxonomies will still be different. Text labels may provide some initial search capability but to improve the integration of data the taxonomies will also need integration. ResearchSpace should allow multiple co-referencing of data from different collection systems (Through a co-reference sub-system similar to that provided by <http://sameas.org/>, tailored to the needs of the project).

7.10.2 A project may have a co-referencing profile at a project level or may allow different users to create their own set of co-referencing queries and have personal profiles. Therefore a project may have many different sets of co-referencing rules.

7.10.3 In addition, different organisations may wish to propose and agree an official co-referencing view. The different co-referencing views above should be created externally from the RDF.

8 ResearchSpace Environment



ResearchSpace System		author: Dominic	created: 05/12/2010 21:01:26	THE BRITISH MUSEUM ³
Collaboration		version: 1.0	modified: 27/05/2011 10:33:38	
		status: created		
		rsflow (3).vsd		

8.1 Environment Processes

- Monitor project – monitoring research projects through the ResearchSpace dashboard.
- Provide support – provide administrative and technical support to registered projects.
- Set research questions – the questions set for the collaborative environments
- Review Research questions – review of progress and assessment for the generation of new questions and avenues of research within a project
- Review Project Configuration – make decisions on the components (research tools used) and their configuration.
- Change Project Configuration – make changes to the project environment.
- Identify additional data – determine whether new data sources are required.
- Request additional data – commission new data from contributing partners.
- Use research tools – make use of the available research tools to analyse and annotate the data.
- Use publication tools – publish data to the shared RDF store and through web applications.
- Receive request for data - consider a request for data from a ResearchSpace project.
- Upload – upload information or data to ResearchSpace.

8.2 Introduction

8.2.1 The home page is both the welcome page for ResearchSpace but also a place where a user can quickly catch up on the activity in the space. The Project Administrator's dashboard should provide the ability to create a project announcement or news item to be displayed on all dashboard welcome pages for registered project users. For each collaboration area the last post should be displayed. There should be a direct link to join the projects collaboration activity. Documents and assets posted since last login should be listed with links.

8.2.2 Messages from other project users should be viewable. A simple user interface will be provided to allow a non-technical user to set these questions. By default, although a social networking component may be included in the project, the Project Administrator is the only person who can initiate a discussion or blog unless otherwise configured. Therefore the Project Administrator can initiate all discussions and questions or allow other members of the team to initiate them also.

8.3 ResearchSpace Administrator's Dashboard

8.3.1 The RS Administrators dashboard is a web application that provides the functionality required for the ResearchSpace administrator to monitor and manage ResearchSpace projects and resources. The aim of the dashboard is to expose all the main functionality required for the administration of ResearchSpace into one intuitive control panel. The dashboard must provide the following functionality;

- **Welcome Page and Project Information:** - The default dashboard display is a welcome page. This view should also provide ResearchSpace information in terms of the latest activity in terms of changes in ResearchSpace assets (e.g. new documents and assets uploaded and transfers from projects)
- **Access to RS Registry:** - Provide a list of projects and summary information about them from the RS Registry. The administrator should be able to see a summary list which can be sorted by lead organisation, principal investigator, project description, status and date of creation. The administrator should also have the option of viewing closed projects. The dashboard should allow the administrator to drill down to see all project information and provide an edit mode to change field values.

- **User Accounts:** - The dashboard should provide full administration of both ResearchSpace and Project user accounts allowing creation and closure of accounts.
- **Access to RS Catalogue:** - The dashboard should provide information about the information and data entering and leaving the ResearchSpace data stores which should be viewable per project and by period. The dashboard should allow administrators to set quotas on the amount of space available to a project and for the ResearchSpace overall.
- **Capacity Statistics:** - Information about the amount of space used overall and in each project.
- **Messaging functionality:** - The ability to view and send messages to ResearchSpace users and Project Administrators
- **Alerts and Audit Information:** - Provide administrator with notification if projects have used all their quota, if a project is due to finish soon, if there has been a security problem, etc.

8.4 Project Administrator's Dashboard

8.4.1 The Project Administrators dashboard is a web application that provides the functionality required for the project administrators to monitor and manage ResearchSpace projects and resources. The aim of the dashboard is to expose all the main functionality required for the administration of a ResearchSpace project into one intuitive control panel. The dashboard must provide the following functionality:

8.4.2 **Welcome page and project Information:** - The default dashboard display is a welcome page which provides the latest announcements and news for the project. This view should also provide project information in terms of the latest activity in collaboration input and changes in project assets (e.g. new documents and assets uploaded) from other team members and user activity on the project at the last two logons together with the dates on those logons. Activity information should provide links to allow the Project Administer to quickly navigate to a previous activity area.

8.4.3 **Project Communication:** - The Project Administrator should be able to create project news and announcements that appear when Project users logon to the environment.

8.4.4 **Access to RS Registry:** - Provide a list of projects and summary information about them from the RS Registry. The Project administrator should be able to see a summary list which can be sorted by lead organisation, principal investigator, project description, status and date of creation. The Project administrator should also have the option of viewing closed (and therefore published) projects. The dashboard should allow the administrator to both drill down to see information for his/her projects, and provide an edit mode to change field values.

8.4.5 **User Accounts:** - The dashboard should provide full administration of Project user accounts allowing creation and closure of accounts.

8.4.6 **Access to RS Catalogue:** - The dashboard should provide information about the information and data entering and leaving the project data stores. It should also be possible to view the status of project data and information that has been published to the ResearchSpace stores which should be viewable by period and/or by user.

8.4.7 **Capacity Statistics:** - Information about the amount of project space available and its usage

- Access policy to the site
- User setup and notification
- Social Networking components
- Workflow decisions

- Document management library setup and configuration
- Asset Management configuration
- Creating the research questions and objectives
- The ability to move assets from Project stores to ResearchSpace stores

8.5 Project Team Member Dashboard

8.5.1 The dashboard is a web application that provides the functionality required for the ResearchSpace Project Team member to manage resources. The aim of the dashboard is to expose all the main functionality required for the use of ResearchSpace within the context of the Project into one intuitive control panel. The dashboard must provide the following functionality:

8.5.2 **Welcome page and Project Information:** - The default dashboard display is a welcome page which provides the latest announcements and news for the project. This view should also provide project information in terms of the latest activity in collaboration input and changes in project assets (e.g. new documents and assets uploaded) from all other team members and user activity on the project at the last two logons of the team member with the dates of those logons. Activity information should provide links to allow the user to quickly navigate to a previous activity area.

8.5.3 **Access to RS Registry:** - Provide a list of projects and summary information about them from the ResearchSpace Registry. The user should be able to see a summary list which can be sorted by lead organisation, principal investigator, project description, status and date of creation. The user should also have the option of viewing closed projects.

8.5.4 **Workflow Configuration:** - The Project User should have the ability to configure the following workflows;

- The ability to determine the status of new data or elements, as determined by the Project Administrator
- User Account: - The dashboard should provide access to the user account profile and allow changes in password and personal and contact details.
- Access to RS Catalogue: - The dashboard should provide information about the information and data entering and leaving the project data stores and on the status of project data and information that has been published to the ResearchSpace stores which should be viewable by period and/or by user.
- View Capacity Statistics: - Information about the amount of project space available and its usage.
 - Access policy to the site
 - User setup and notification
 - Social Networking components
 - Workflow Configuration
 - Document management library setup and configuration
 - Asset Management configuration
 - Creating the research questions and objectives.
 - The ability to move assets from Project stores to ResearchSpace stores.

8.6 Workflow and Content changes

8.6.1 ResearchSpace seeks to make the working environment as fluid and collaborative as possible. As such formal workflow will be kept to a minimum. However, it is expected that a project will generate a large number of documents and annotations. This body of work may be difficult to keep track of and when it comes to selecting the final data to be presented on a project site, the process of going back

through the work and selecting the final project view may be difficult. Workflow will therefore mostly be orientated around content alerts and selection. The workflow system should therefore support the following:

8.6.2 **Data Annotations:** - When a data annotation is made by a member of the team this activity is logged and becomes part of the Project Administrator's activity list. The Project Administrator can then assign a status to the content which will also provide a link to the area of ResearchSpace that the annotation was made to allow the administrator to check context.

8.6.3 Image Annotations: - As above

8.6.4 **File Submissions:** - When a document is loaded into the project or ResearchSpace asset libraries then this activity is recorded and is listed in the Administrator's activity list. The Administrator can then, from the activity list, open the document for review. The status flags should be configurable for each project (and should be multiple) but by default are;

- **New** – The data has been uploaded or a new data record created and requires review before acceptance into the Project or before conversion.
- **Public** - The information can be release to the ResearchSpace shared repositories at any time.
- **Release on Project End** - The information can be released at the end of the project
- **Release during project on authority** – The information requires authority from the data owner before release.
- **Project Only** - (not for release) – The information will not be released to the shared Repositories
- **Source for RDF data conversion only** – The asset is a data source and not for use as an asset in the project and will not be shared.
- **Converted** -The document has been converted into structured data.
- **Project Approved** – The asset or data is approved for project publication.
- **Approve Converted Document** – The document has been converted to structured data and awaits approval.
- **Superseded by** (another document) – The document or asset has been replaced by a different document.
- **Withdrawn** – Denotes that the asset has been withdrawn from the ResearchSpace system.
- **Deprecated** – The data has been depreciated by another version.
- **Sequence number** – Indicates the preference in terms of which assets to display on web publication

8.7 Workflow and Data

Data will be uploaded to ResearchSpace either in a format that requires conversion to the ResearchSpace CRM structure or in a format already in the required structure. The former will be uploaded as files in a delimited, or XML format and should be provided with an appropriate status – 'New'. New data will need analysis and mapping to the correct format for storage in ResearchSpace. Data uploaded using the ResearchSpace conversion features will have the status of – 'New' and 'Converted'. This should provide the opportunity to review the conversion before a project status is supplied. The data input research tool should be user configurable to newly converted data.

8.8 ResearchSpace Workflow – Configuration

8.8.1 The project administrator by default should have rights to change the status of information or data. However, the Project Administrator can also delegate, through the dashboard, this right to other

members of the team. In addition, the project administrator can configure what alerts should be triggered and to what users the alerts should be directed.

Areas in which workflow may apply are;

- Classifying or verifying the status of data and files.
- Alerts triggered by annotations and data being added through research processes.

8.8.2 The other form of work is for validating the quality of data conversions and data input. Where information is converted into structured formats that comply with the ResearchSpace data templates or when data is inputted into the online data entry screens there is a need to ensure that the data has been supplied in the correct format.

8.8.3 **Data Entry Form** - The Project Administrator should be able to configure the data entry screen(s) for the project by selecting the fields from the ResearchSpace template. The Project Administrator can also indicate the fields that are mandatory. ResearchSpace should develop both general and specific data validation (as required) to ensure that manual data entry is correctly entered. Data entered on the Data Entry form is set by default to 'Public' in the RS Catalogue but can be set to, 'Release on Project End' or, 'Project Only'.

8.8.4 **Thesauri and Authority Files** - Some fields will require the use of authority files and thesauri. The data entry form should provide a thesauri control which can be populated using the SKOS thesauri that have been imported into ResearchSpace. ResearchSpace will by default offer a set of thesauri to which other thesauri have been co-referenced. However, the Project Administrator should be able to, from the dashboard, configure the appropriate SKOS thesauri to be used with their project's data entry scheme.

8.9 Navigation

8.9.1 ResearchSpace should have a consistent design and navigation system. The navigation system should provide a fluid relationship between the different ResearchSpace components, Research Tools, Collaboration Tools and data stores. The collaboration tools must have navigation that allows the launch of research tools directly from collaboration text editors, and allow a return to that editor with session state intact regardless of whether other research tools have been launched in the mean time.

8.9.2 This is particularly important as collaboration content can be used in research tools and data found by using research tools can be returned into a collaboration editor using the ResearchSpace, 'data basket'. The act of exchanging data from one environment to the other should create links that provide navigation from the collaboration tools to the research tools. For example, data found using a research tool and brought back for use in a forum should create a link that would launch a research tool using the link as a parameter.

8.9.3 **ResearchSpace Toolbar**: - The toolbar will be available throughout the ResearchSpace environment. It will be added to all collaboration editors allowing the user to launch available research tools, access project document and asset libraries and access their dashboards. This should be done without losing the state of the collaboration editor.

8.10 Collaboration Environments

8.10.1 It is currently envisaged that the collaboration environments will be supplied with the chosen Content

Management System. ResearchSpace can theoretically operate within any collaboration environment, Blog, Forum, Wiki etc. The ResearchSpace toolbar should be available in the editors of these collaboration tools. This means that ResearchSpace tools can be launched during the process of creating a post or contribution. The following additional features are required:

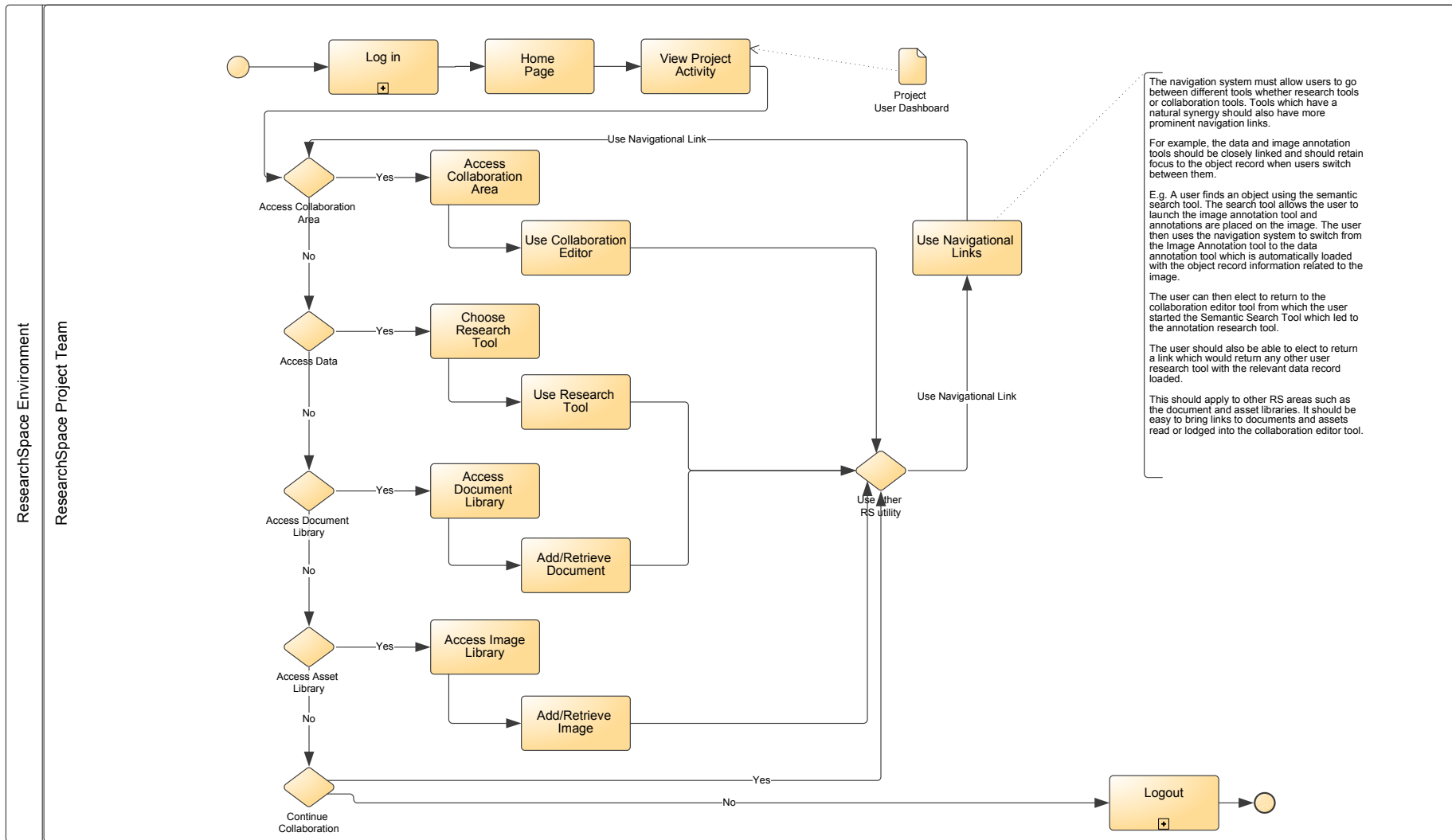
- 8.10.2 The state of the editor should be preserved so that a user can launch a research tool (and launch others) and return to the editor in the state that it was left. The user should be able to bring back data and links to the collaboration editor and insert these into the post for other users to pick up.
- 8.10.3 It should be possible to view collaboration both in terms of research tool contribution and contextual information. This means that data entered from a research tool, launched from a collaboration editor, should be viewable separately from the additional narrative inputted into the post. This view will show in date sequence the annotations and other data inputs that occurred during collaboration against a particular research objective. It should also be possible to filter this by contributor.

8.11 Researcher's Data Basket

- 8.11.1 While using the research tools users will come across data that they wish to use in the collaboration environments. It should be possible to add selected data or links to data or assets into a 'basket' which can be used to insert into a collaboration post. This should include:
 - Data – a user may select some data or an annotation on data.
 - Links to data – a link that when another user clicks on it takes them to the required data.
 - Links to images – as an image tag and a hypertext tag.
 - Links to searches - the ability to provide a link that invokes a search. For example, a user searches and decides to save the search for another to use.

Data should be specifically formatted and tagged with its RDF identifier so that it can be reported and used for a search.

9 ResearchSpace Research Tools



The navigation system must allow users to go between different tools whether research tools or collaboration tools. Tools which have a natural synergy should also have more prominent navigation links.

For example, the data and image annotation tools should be closely linked and should retain focus to the object record when users switch between them.

E.g. A user finds an object using the semantic search tool. The search tool allows the user to launch the image annotation tool and annotations are placed on the image. The user then uses the navigation system to switch from the Image Annotation tool to the data annotation tool which is automatically loaded with the object record information related to the image.

The user can then elect to return to the collaboration editor tool from which the user started the Semantic Search Tool which led to the annotation research tool.

The user should also be able to elect to return a link which would return any other user research data with the relevant data record loaded.

This should apply to other RS areas such as the document and asset libraries. It should be easy to bring links to documents and assets read or lodged into the collaboration editor tool.

ResearchSpace System	author: Dominic	created: 05/12/2010 21:01:26	THE BRITISH MUSEUM
Use Research Tools	version: 1.0	modified: 27/05/2011 10:33:38	
	status: created		
rsflow (3).vsd			

9.1 Process for using Collaboration and Research Tools

- Login – the process and rules for logon to ResearchSpace
- View project activity – through the use of the ResearchSpace dashboards
- Access collaboration editor – linking from the dashboard to the particular collaboration area required
- Use collaboration editors – the use of the particular collaboration editor for posting information
- Choose research tool – from the tools that have been selected for the project
- Use research tool – the use of the research tool launched from the collaboration editor
- Access document library – access to the RS document libraries
- Add, retrieve, edit document – use of the RS document libraries
- Access digital asset library - access to the RS digital assets libraries
- Add, retrieve, edit digital asset – use of the RS digital assets libraries
- Use navigational links – Navigate between the different RS tools
- Logout – Logout of the system

9.2 Summary

9.2.1 **Semantic Search Tool** – The semantic search tool will provide the main mechanism for finding and accessing ResearchSpace data and assets. It is envisaged that this tool will provide a number of different mechanisms for searching the ResearchSpace repository including the ability to specify rules used for inference and display relationships to other data based on initial results.

9.2.2 **Data Input Tool** – This will allow standard data creation, modification and deletion based on data validation rules and taxonomies. [Within same tool as Data Annotation]

9.2.3 **Data Annotation** – The ability to add additional information using a controlled annotation ontology or with free text.

9.2.4 **Image Annotation** – The ability for researchers to annotate an image either against the whole asset or against a particular point or region of the image. The data would be stored as RDF and be searchable through the semantic search tool. Annotations should be both controlled using the CRM (Conceptual reference Model) framework and also allow the option for uncontrolled comment. Note: <http://www.cidoc-crm.org/docs/fin-paper.pdf> & <http://www.w3.org/2005/Incubator/mmsem/XGR-vocabularies/#existing-SI> & <http://eprints.aktors.org/425/01/OntoMedia.pdf>

9.2.5 **Image Zoom (and Annotation)** – Many ResearchSpace users will want to upload large high resolution images and allows colleagues the opportunity to zoom in and thus examine images in detail. The ability to annotate at different levels of zoom would also be useful. In the event the annotations should be available to the user and take him or her to the exact point where the annotation was made. [Within same tool as Image Annotation].

9.2.6 Note on Annotation – The ResearchSpace annotation requirement for both data and digital assets is a strong one. The stage 2 specification work will look carefully at existing work in this area and particularly the Open Annotation Collaboration project (www.openannotation.org) which is developing an approach to RDF annotation. The specification will be used for the working prototype.

9.2.7 **Image Compare** – The ability to compare different images in order to uncover areas of interest for annotation. This may involve operations such as;

- Scaling two different images so that they fit one over another
- The ability to compare images through transparency or pixel difference

- It would also be assumed that basic image editing features would also be available
- 9.2.8 **Relationship / Link Editor** (internal and external) – The ability, through controlled terms, to link object records and group these relationships so that a defined story or pathway can be recorded for others to follow.
- 9.2.9 **Version Comparison** (Track Changes) – It should be possible to see all the different versions of a data field and see who created it.
- 9.2.10 **Geographical Mapping** – It should be possible to map data objects according to different types of recorded location, whether that is a production place, a place of birth, the origin of a material and so on. The user should be able to define the scope and scale of the map and the data elements that should be plotted.
- 9.2.11 **Timeline Mapping** – The ability to map objects against a graphical timeline. It should be possible to have different modes for recording frequency as well as individual object mapping. For example, it should be possible for a researcher to plot a large number of objects on a timeline to show graphically information such as the number of particular objects found as a function of time.
- 9.2.12 The ResearchSpace Requirements RS Catalogue also implies a number of other tools for collaboration, such as video conferencing and online educational tools.
- 9.2.13 In addition, ResearchSpace should support the inclusion of other systems for activities such as real time collaborative editing, video conferencing and other communication tools.

10 Research Tools – Detail

10.1 Introduction

10.1.1 Research Tools should be created with both researchers and end users in mind. Appropriate administration and configuration interfaces should be provided to control styling, access rights and the available functionality, to allow tools to act as read only tools where appropriate, allow certain functionality to be removed if potentially not appropriate for the published site.

10.2 Semantic Search Tool

10.2.1 The search tool should provide different searching techniques. Although these are described separately the search tool could combine different types of search within the same user interface and allow users to switch between different methods and views of the data. Importantly, users should be able to browse through data and be provided with different avenues of investigation. They should also be able to change criteria and filters as they go with an emphasis on exploration. The wire frames below are examples to illustrate particular search approaches. Search may require specific controls or filters for different types of records which may include objects, sites and buildings. Additional searching functional requirements and wire frames will be provided.

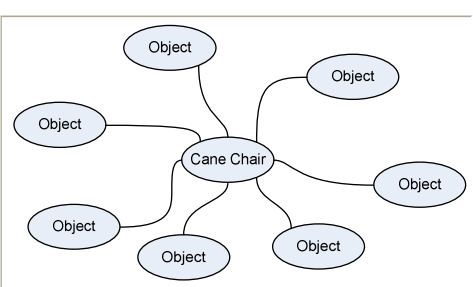
10.2.2 **Ontology and Taxonomy Searching** – All data in the RDF repository is organised using the CIDOC-CRM framework. In doing this a level of harmonisation is already achieved which can be utilised by a search mechanism. The ontology search would allow searching by CRM ontology terms such: what, where, who and when. Out of the box it would be envisaged that some cross store hits would be possible using label text within the RDF. However, full cross data ontological searching would require a co-reference of terms using by the different organizations submitting data. To aid this process ResearchSpace aims to provide a non-technical tool for co-referencing terms.

10.2.3 **Data Relationships** – The search environment should suggest different avenues of exploration by showing related results using different categories. For example, a particular result may have many different connections with other objects which may be of a similar object type but may also have connections with people, artists, writers etc. The connection may be an indirect relationship (for example based on a style of painting or an event or meeting or because of membership to a particular school.) Another avenue might be for objects that originated from the same or connected geographical areas or collections. Scholars using the search tool should be presented with a wide range of potential connections and therefore different avenues of exploration and discover.

10.2.4 **Scoped Searching** – This would allow scholars to concentrate on searching material on a particular object or subject and therefore narrow down the potential results from the outset. However, it should also be possible to view related material outside the particular sphere to provide a view providing contextual information. For example, a scholar may be interested in searching coins and concentrate on these results, but still might also be interested in seeing associated information about containers, manufacturers or other associated finds.

10.2.5 **External Data Sources** – The potential relationships to be discovered should extend to data sources outside the ResearchSpace sphere. The user should be able to search other well known sources of information including, for example, Dbpedia. The system should allow for the expansion of these different data sources which can be chosen within the search environment including web sites that use the RDFa standard.

Explore



Graphical representation of the selected object with connections. Ability to select connecting object and see its connections.

<< < 1 of 5 > >>

Results

Result 1 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc....

Result 2 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc....

Result 3 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc....

Summary results

Culture	▼	Schools	▼
Ethnic Group	▼	Styles	▼
Escapement	▼	Subject	▼
People	▼	Material	▼
Places	▼	Technique	▼
Object Type	▲	Ware	▼

Canop

- CANOPTIC COFFIN
- CANOPTIC JAR
- CANOPTIC URN
- CANOPTY

- ☞ CANOPTIC JAR
- ☞ CANOPTIC URN
- ☞ URN
- ☞ CANOPY
- ☞ CANTARO

Add

SCOPE NOTE

Result 1 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc....

Expanding terminology search boxes with suggestions as you type.	▼
--	---

Search Terms

- ☞ CANOPTIC JAR
- ☞ CANOPTIC URN
- ☞ URN

Remove

Added search terms with complete hierarchy. Ability to select individual terms or remove

Details

Object types
canopic urn (scope note | all objects)

Materials
terracotta (scope note | all objects)

Techniques
modelled (scope note | all objects)

Production place
Made in Italy (all objects)

Place (findspot)
Excavated/Findspot Chiusi (all objects)

Date
625BC-600BC (circa)

Period/Culture
Etruscan (scope note | all objects)

Description
Terracotta canopic urn set upon a chair, the lid modelled as a head with holes probably for attachment of a mask.

Dimensions
Height: 18 inches
Height: 23 inches (with chair)

Curator's comments
The type of vessel is so called because of its resemblance to Egyptian canopic jars.


Edit / Annotate data

Annotate Image

Details pane for the object, site, building

Coins

Results

Result 1 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Result 2 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Result 3 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Result 1 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Result 2 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Result 3 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	

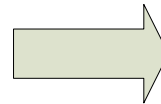
Relationships



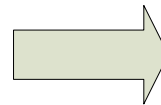
Cultural



People



Object Type



Sites

Relationship Data Source

- ResearchSpace
- DbPedia
- Freebase
- Europeana

Cultural Result 2 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Cultural Result 2 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Cultural Result 3 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Cultural Result 4 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Cultural Result 4 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	
Cultural Result 5 – Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut et venenatis metus. Proin consequat, risus a blandit tempus, ipsum elit luctus tortor, non aliquam enim libero et orc.....	

- 10.2.6 **Keyword Searching:** – The ability to search structured data using an indexed keyword search. The system would also suggest variations of the keyword based on variations within the various taxonomies to widen the search further.
- 10.2.7 **Taxonomy Searching:** – This would provide a dynamic search interface using the data categories to home in on a result with the ability to filter further using taxonomy terms. This might work in a similar way to the Finnish Museum’s prototype, e.g. (1) use the location taxonomy to filter my search down to the UK; (2) switch to the materials taxonomy and view what material terms had results at this level of the search (3) then select a term to filter down further (4) consult the available results against the people taxonomy to locate a particular artist.



1 <http://www.museosuomi.fi/>

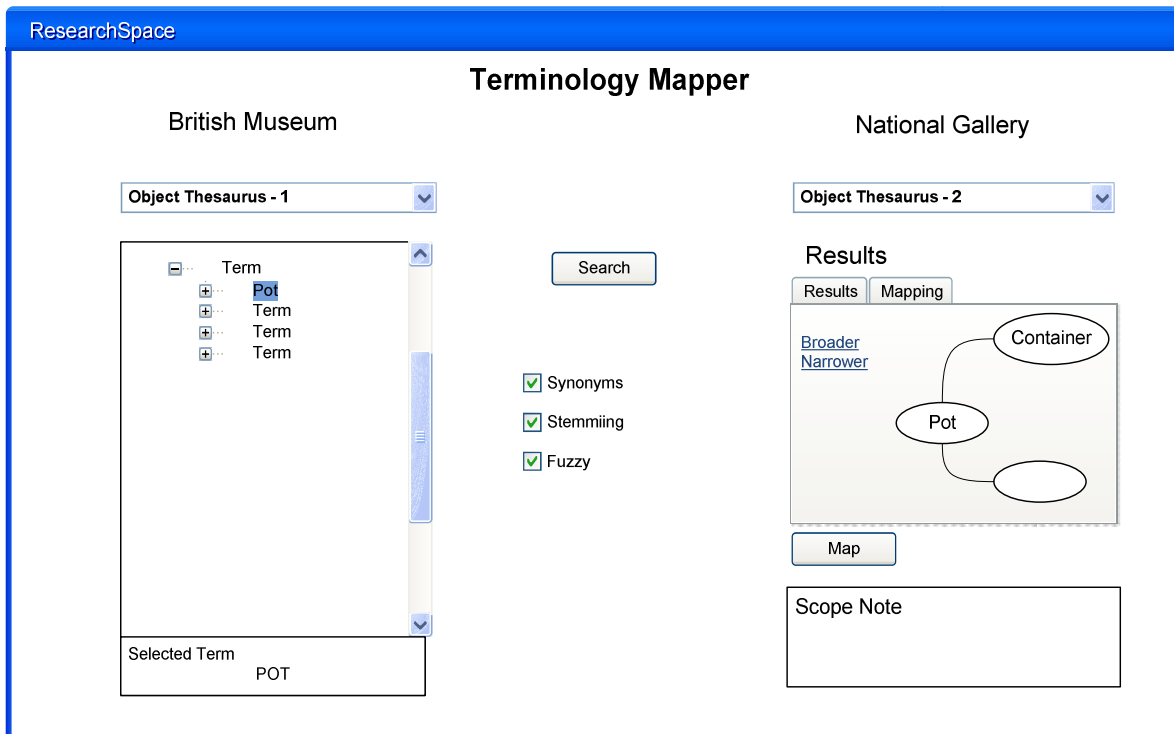
- 10.2.8 **Relationship Finder** – The search system should also provide an interface that allows the researcher to make use of all the potential relationships for a given object potentially across all the data in the ResearchSpace repositories and move through those connections and relationships. In effect this provides a way of building pathways through the data and of uncovering relationships that would not be discovered ordinarily using standard database queries. The researcher should also be able to change the behaviour of the relationship finder by asserting rules. For example:
- If data about a sword made in the 12th century recorded a particular steel composition and research showed that this particular composition and the particular period indicated a place of production then this assertion could be made across all datasets in order to uncover more results and relationships.

10.2.9 **Free Text Searching** – Research will require an indexing system to support text searching.

10.3 Terminology Mapper

10.3.1 A data and graphical tool to allow terms used in two different authority files to be mapped. The user would select two sets of vocabulary (two thesauruses for example) that have been ingested into the ResearchSpace CRM/SKOS format. These would be loaded into the tool and allow terms to be compared both manually and through searching initiated by choosing a term in the source thesaurus and using it as the search criteria for locating candidate terms within the target vocabulary. The user should be able to map single or multiple terms.

10.3.2 The tool would then allow the user to select a term in one thesaurus to initiate a search in the other. This search would employ techniques such as;



10.3.3 Word Stemming

- Synonyms
- Fuzzy Matching

10.3.4 The result would be ranked appropriately and display matched terms and their scope notes. The user should be able to select a term from the results to display the structure (broader and narrower terms) around it so that a final mapping decision could be made. Is it possible that the term could be mapped against multiple terms depending upon how the two thesauri compare in detail and precision.

10.3.5 The system would then allow the user to correct or complete the mapping and then save it as a mapping profile. When mapping profiles are saved this should be alerted in the Project Administrator's dashboard as a new project profile. It should also be possible for different mappings between the same vocabularies to be saved and used as alternatives.

10.4 Data Annotation Tool

10.4.1 Object records that are uploaded to ResearchSpace may be annotated in three ways.

- Researchers may wish to add data that is currently missing from a record. The missing data may be uncovered as a result of the research collaboration.
- Researchers may wish to propose a replacement of existing data. For example, the project may conclude that existing data is inaccurate or there is some doubt and wish to propose an alternative.
- Researchers can add new fields with an appropriate data type. These new fields may be part of the ResearchSpace data templates and have CRM mappings but may also be outside this structure. Fields outside the ResearchSpace structure should be recorded using a standard mechanism but be alerted to the Project Administrator and ResearchSpace administrator for incorporation into the ResearchSpace CRM data structure. By default, no new data fields should be allowed without this being configured by the Project Administrator whose dashboard will control this setting.
- Researchers wish to annotate data with other data which is not part of the object data structure but has some relevance to the research being conducted and would provide information that would increase understanding or help with the research. This could be some contextual information or document a relationship to other data in the repository. For example, there may be an interesting connection with another object or some qualification may be required to make better sense of the data that cannot be entered into a data field itself.

10.4.2 To annotate data on a single record the researcher would first find the record. Once the record has been found a form would provide the data already available. The form should have configuration settings so that a researcher can see various views of the record with different versions of annotation. For example, the researcher may wish to see the original uploaded data only, or the latest version of the data or all the versions of the data.

10.4.3 Clicking on a particular field will open up a new edit box. The researcher will indicate whether s/he is adding new structured data (in which case the edit box will validate the entry) or an additional annotation. It is possible that these additional annotations could be qualified by a given set of predicates

10.4.4 Data entry against the object data will create a new version of the data for the object record. This information will be attributed to the researcher and the project. The date of the version should also be recorded. The researcher should also be given the opportunity to annotate the entry – perhaps to document the rationale for the data entry.

10.4.5 For example, a researcher disagrees with a place of production for a particular object. A new version of the place of production is entered and a reason for this is added as an annotation. The new version itself could be also predicated, e.g. “proposed”, “proven” etc. It should also be possible to predicate additional non-object record annotations.

10.4.6 Annotations which seek to establish a relationship between data in two different object records will also require the ability for two records to be selected and arranged on the screen so that a researcher can establish a link – perhaps by dragging the mouse from one data element in one record to another data element within another record and allowing the researcher to enter the reason for the relationship.

10.4.7 The annotation should also collect the following information taken from W3C standard annotation class.

10.4.8 **annotates** - [<http://www.w3.org/2000/10/annotation-ns#annotates>] - Relates an Annotation to the resource to which the Annotation applies. The inverse relation is 'has Annotation'. Defined by [<http://www.w3.org/2000/10/annotation-ns#>]

- **author** - [<http://www.w3.org/2000/10/annotation-ns#author>]- The name of the person/s or organization most responsible for creating the Annotation. Sub property of [<http://purl.org/dc/elements/1.1/creator>] - Defined by – [<http://www.w3.org/2000/10/annotation-ns#>]
- **body** - [<http://www.w3.org/2000/10/annotation-ns#body>] - Relates the resource representing the 'content' of an Annotation to the Annotation resource. Sub property of related - [<http://www.w3.org/2000/10/annotation-ns#related>] - Defined by [<http://www.w3.org/2000/10/annotation-ns#>]
- **context** - [<http://www.w3.org/2000/10/annotation-ns#context>] - The context within the resource named in 'annotates' to which the Annotation most directly applies. Defined by [<http://www.w3.org/2000/10/annotation-ns#>]
- **created** - [<http://www.w3.org/2000/10/annotation-ns#created>] - The date and time on which the Annotation was created. yyyy-mm-ddThh:mm:ssZ format recommended. Sub property of [<http://purl.org/dc/elements/1.1/date>]. Defined by [<http://www.w3.org/2000/10/annotation-ns#>]
- **modified** - [<http://www.w3.org/2000/10/annotation-ns#modified>] - The date and time on which the Annotation was modified. yyyy-mm-ddThh:mm:ssZ format recommended. Sub property of [<http://purl.org/dc/elements/1.1/date>]. Defined by [<http://www.w3.org/2000/10/annotation-ns#>]
- **related** - [<http://www.w3.org/2000/10/annotation-ns#related>] - A relationship between an annotation and additional resources that is less specific than 'body'. The 'related' property is expected to be subclassed by more specific relationships. Defined by [<http://www.w3.org/2000/10/annotation-ns#>].

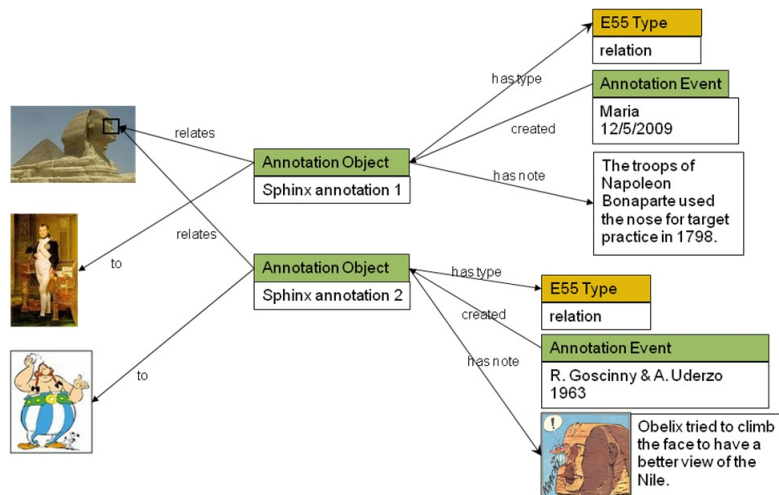
10.5 Image Annotation Tool

10.5.1 The image annotation functionality of ResearchSpace should address three different types of annotation operation:

- The ability to apply a text annotation against a region of an image. The region should be one defined by the user. It should also be possible to apply text to more than one region of the same image.
- The ability to annotate a relationship between regions on two different images.
- The ability to annotate a relationship between a region on an image and data within an object record.

10.5.2 This means that the tool should have the ability to display two object records on a single form at the same time. It is envisaged that the two records would be placed side by side, with the ability to switch on each side between a data view and an image view. As in the data annotation tool the researcher would have the ability to draw links between an image region and a data field or between two regions.

10.5.3 Also, it should be possible to provide configurable controlled predicates for the annotations.



From www.3dcoform.org

10.6 Image Zoom (and annotation)

10.6.1 ResearchSpace requirements include the use of high resolution images. In an online environment these images can practically only be provided using streaming technology. Currently, open source software such as IIP Image, provide the ability to deliver high resolution images by tiling different levels of zoom and delivering these to the browser when required rather than serving the whole image in one go.

10.6.2 Annotations within images should be clearly displayed next to the image and clicking on the annotation will display the region of the image that the annotation refers to. In the case of annotating high resolution images the ability to annotate at different levels of zoom would be highly desirable. In this case a researcher could zoom into the image to the required degree and embed the annotation at that level. Clicking on the annotation listed by the side of the image would cause the correct tile to be served so that others can view the annotation at the level of zoom at which it was created.

10.7 Image Compare

10.7.1 For conservators and scientist the ability to make use of different images created through the use of scientific equipment such as x-ray (radiographs= images) and infra-red (reflectograms= images) allows annotations on the condition or an object or reveals information about the object that cannot be seen by the naked eye but which can hold important information about the history, techniques and provenance of the work. It is usually necessary to closely compare visual/X-ray and IR images together to better understand and interpret the information which each contains. It would therefore be useful to provide these annotations with reference to a normal representation of the object which can reinforce scientific findings further.

10.7.2 The image compare tool therefore requires that images can be manipulated such that different types of images can be overlaid and that annotations can be applied on this consolidated image. This means it must be possible to scale the images appropriately (compensate for differences in image size) and for layered images, to adjust settings to provide the right level of opacity. It should also be possible to provide a simple side by side comparison which can also be annotated and reviewed by others at a later date.

10.8 Relationship / Link Editor

10.8.1 Relationships and links between data should be supported both within ResearchSpace and also to

other sources of information and data outside ResearchSpace. The following are envisaged:

- A hypertext link can be made from an object record to an external web site page.
- A hypertext link as an annotation to a piece of data to an external web site page.
- An object to object RDF association within ResearchSpace.
- An object RDF association to a URI outside ResearchSpace.
- Links should also be supported by the ability to add further contextual annotation.

10.9 Version Comparison (Track Changes)

10.9.1 ResearchSpace has the opportunity of bring together a whole body of work on a particular subject. As time goes by more information will be added and in RDF data is always added but not taken away.

10.10 Geographical and Timeline Mapping

10.10.1 This system tool is required for mapping objects or events from the various imported datasets. One issue is that different datasets uploaded by different organisations will use different terminology for geographical names and may not contain geographical coordinates. The ResearchSpace System should provide a tool for converting geographical names into coordinates and for generally supporting longitude and latitude coordinates information.

10.10.2 Researchers should be able to define a set of data which contains location information and represent the subject of this data on the map. This may be the current location of an object, its find spot and the place of production, for example. It may also be the location of an event which may span a particular period of time which should also be capable of representation. The data field defined could also be an annotation placed on the data by the researcher. It should therefore be possible to determine a time frame for any particular map and for information to be exposed that provide date and location context.

10.10.3 In order for the tool to provide a semantic element it should be possible to populate the map based on a set of data and for potential relationship to also be represented within a map and timeline.

worldtimelines.org.uk

Home | Index | Museums | Help | Search | Go

Compare timelines

AD 635 AD 640 AD 645 AD 650 AD 655 AD 660 AD 665 AD 670 AD 675 AD 680 AD 685 AD 690 AD 695 AD 700 AD 705 AD 710 AD 715 AD 720 AD 725

Central England

- Anglo-Saxon women: Penda allies in Cadwallon, ...
- Early Christian art in central England
- Early Anglo-Saxon jewellery
- The kingdom of Mercia

AD 655 Abbey founded at Peterborough

AD 653 Monks AD 659 Wulfhere becomes king of Mercia

AD 655 Death of Penda

AD 676 Ethelred founds a monastery at Bredon

AD 679 Mercians defeat the Northumbrians at the

AD 716 Ethelbald becomes king of Mercia


AD 718 Colwold becomes king of Mercia

AD 709 Cenred becomes king of Mercia

Timeline view

World | British Isles | England

Central England AD 410-1066 Early medieval



The history of central England from the 5th century AD, when Anglo-Saxons were invading the east coast and many Britons fled to Wales, is obscure until the reign of Penda (632-55), king of Mercia. By the early 7th century, Mercia's territory reached from Wales to the Lincolnshire coast, and later even extended as far south as London. Under strong monarchs such as Ethelbald (716-57) and Offa (757-96) it dominated England south of the River Humber into the 9th century.

The Old English name for Mercia was *Mierce*, meaning the people of the borders. Its most vulnerable borders were with Wales to the west and Northumbria to the north. For much of the period central England was a theatre for warfare. The great defence of Offa's Dyke (probably mainly constructed during Offa's reign) kept the Welsh at bay, but there was fighting with the Northumbrians throughout the 7th century, until the Mercians defeated them at the River Trent in 679.

Mercia itself was defeated by Wessex in 825, in a dispute over the 'debatable lands' on its southern border, and lost its overlordship of London and the south. After the Danish invasions of 865, Mercia was divided. The northern half was occupied by the Danes and only a small area in the south remained under Anglo-Saxon rule.

© 2007 The British Museum

11 ResearchSpace Publication

11.1 Templates and Plug-ins

11.1.1 ResearchSpace should provide a standard set of configurable (styling and functions) templates and plug-ins used to create web applications. These should include the following standard templates and plug-ins. These may be based on templates and plug-ins already supplied with the Content Management System.

11.1.2 **Standard Templates:** - For either horizontal or vertical hierarchical navigation menus or a combination of different column and header and footer sections.

11.1.3 **Image Light Box:** - For showing a light box of images based on database query or session selection. This should link to full object records

11.1.4 **Record Display:** - The ability to display fields related to a particular object, site, building, etc. The plug-in should allow the project to configure what fields are shown, in what order and alter the screen space required for the information. The plug-in should allow the application of different style sheets and be configured as the summary results page for a search and for use as the detailed results page for individual records.

11.2 ResearchSpace Tools as public web site plug-ins

11.2.1 Research tools should also be offered as web site tools or plug-ins for general internet users of the site but in a form that can be locked down through configuration, if required, to restrict access (for example read rather than write access. This includes the following:

11.2.2 **Semantic Search Tool** – This should also be available as a user search tool but the project should be able to configure it such that only selected search interfaces (which could mean all) are available.

11.2.3 **Geographical Mapping Tool** – Users should be able to select records using a basket mechanism and choose a geographical field or annotation which than then be represented on a map.

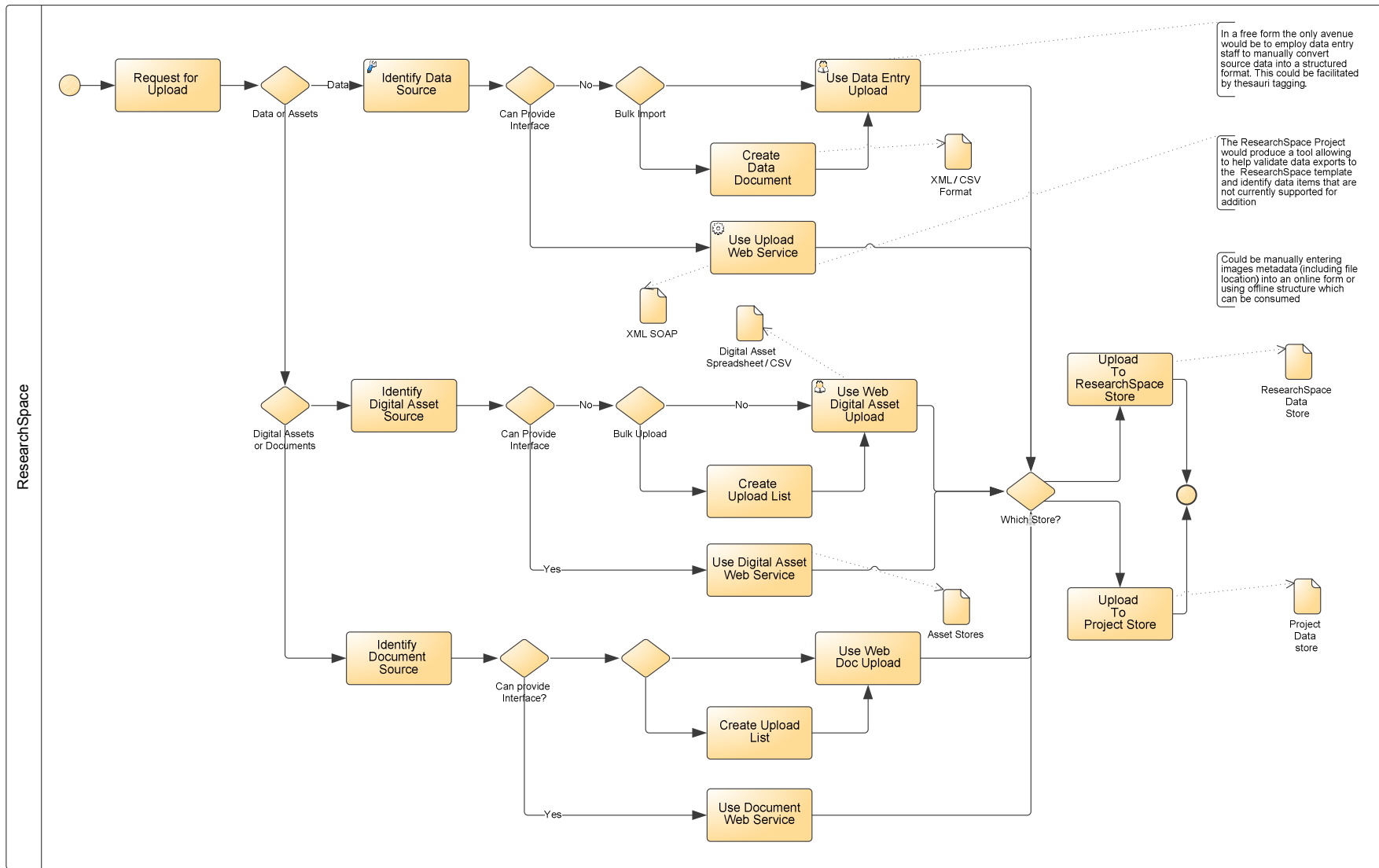
11.2.4 **Timeline Tool** - Users should be able to select records using a basket mechanism and choose a date field or annotation which than then be represented on a map.

11.2.5 **Annotation Tool** – Web sites should provide the ability for public annotations with appropriate safeguards in terms of moderation and identification.

11.3 Security

11.3.1 Although web sites are designed for public publication they should have a security mechanism that supports both anonymous access and specified user access. This is designed to help with testing and review.

12 ResearchSpace Data Gateway



In a free form the only avenue would be to employ data entry staff to manually convert source data into a structured format. This could be facilitated by thesauri tagging.

The ResearchSpace Project would produce a tool allowing to help validate data exports to the ResearchSpace template and identify data items that are not currently supported for addition

Could be manually entering images metadata (including file location) into an online form or using offline structure which can be consumed

ResearchSpace Data Store

Project Data store

ResearchSpace System		author: Dominic	created: 05/12/2010 21:01:26	THE BRITISH MUSEUM
Upload		version: 1.0	modified: 27/05/2011 10:33:38	
		status: created		
		rsflow (3).vsd		

12.1 Introduction

12.1.1 Projects can use only data already existing in the ResearchSpace repositories and data and information uploaded by the other organisations within Projects. Contributing organisations may be part of a research project or simply passive data providers. ResearchSpace will provide the tools and mechanism by which data and information can be uploaded into ResearchSpace and provides the option of allowing conversions to the RDF at the point of upload or will support conversion within the environment. In addition ResearchSpace will support the export of data in a structured XML format so that data can be returned to organisations in a structured format.

12.2 Conversion Templates

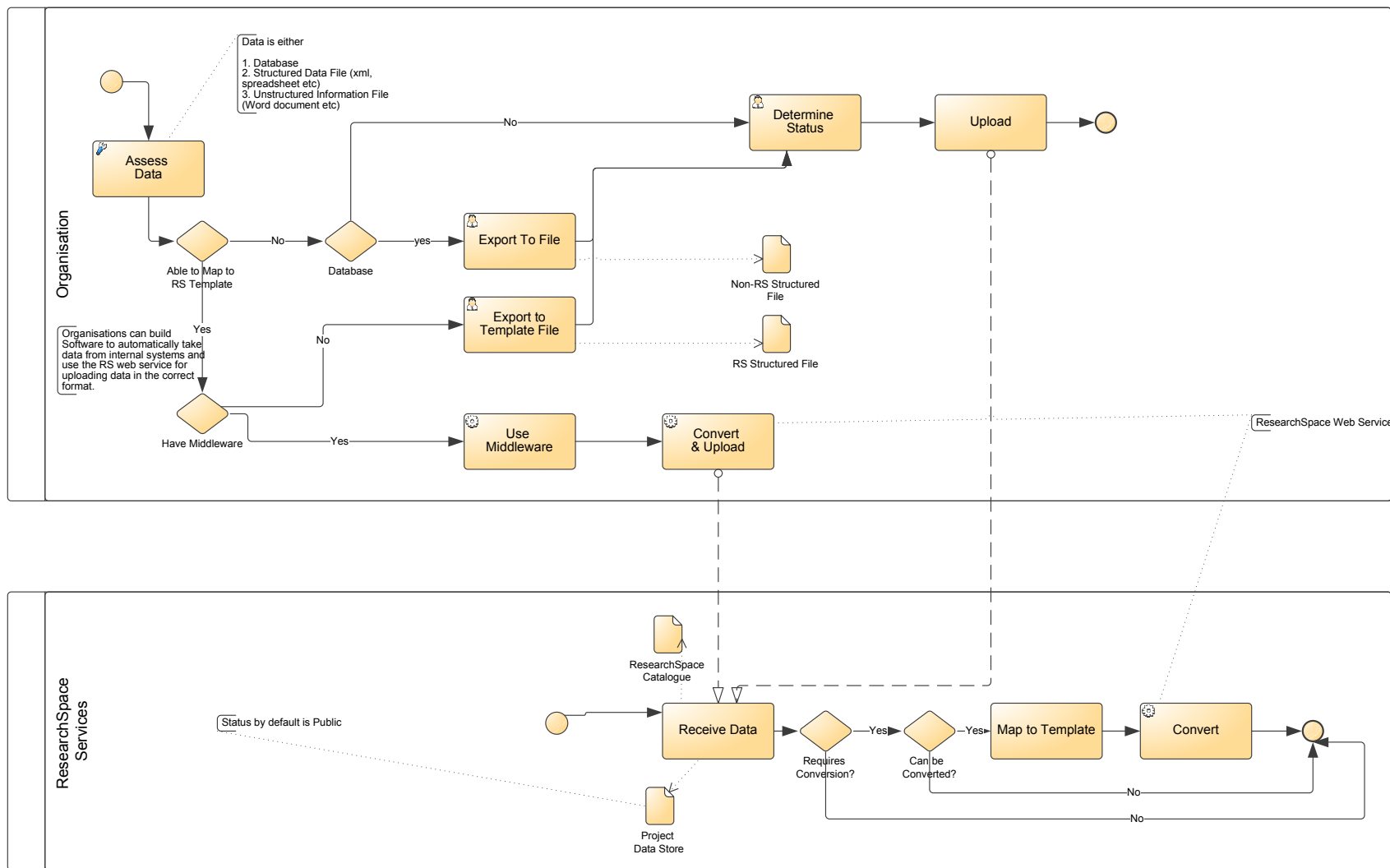
12.2.1 ResearchSpace templates provide a list of data fields that have been mapped to CIDOC-CRM. The CRM is a high level ontology that supports a wide range of different source schemas relevant to cultural heritage. These templates will develop over time as organizations use ResearchSpace and as a result expand the number of data concepts that are mapped and can then be used by others. These templates are expected as a minimum to cover the areas of collections, conservation and science, archaeology and buildings.

12.2.2 When an organisation wishes to contribute structured data to ResearchSpace it has the option of completing a mapping to one or more ResearchSpace templates to the data before upload or developing an interface between ResearchSpace and the internal data source. In both cases the upload web service would validate the data and, if submitted in the correct format, ingest the data in CIDOC-CRM format.

12.2.3 The process and service would be similar for unconverted data already within the ResearchSpace environment. For other unstructured files the opportunity to convert information into CRM structured data would also be possible but would require additional work and tools.

12.2.4 The ResearchSpace project should simplify the process of converting source data into CIDOC-CRM RDF. This would ideally be done through a process of: export of data to a basic delimited format, the use of automation tools to provide an approximate map to CRM entities, tools to support the completion of mapping, and an import process to bring the data into the RDF store.

12.2.5 It is currently envisaged that a CSV format should be specified for the initial export which should cover the capability of the majority of collection management systems. A software tool should be created to provide a non-technical mapping environment and provide an approximate mapping which is gradually developed and refined based on previous mapping exercises. This would include the mapping established as part of the British Museum's data conversion.



ResearchSpace System		author: Dominic	created: 05/12/2010 21:01:26	THE BRITISH MUSEUM
Conversion (1)		version: 1.0	modified: 27/05/2011 10:33:38	
		status: created		
		rsflow (3).vsd		6

12.3 Online Data Entry

- 12.3.1 Contributions of data and assets should also be supported through an online data entry form configured to meet individual project requirements and accessible using a project user account. This form should be outside the ResearchSpace environment and be available for all project members as well as other data contributors who are not part of the project team. This means that user accounts will be required which only provide access to the online form and asset upload facility described below.
- 12.3.2 The form would be designed for organisations without an in-house collection system or appropriate information systems and for general data updates. The form should be generated through a wizard allowing the project to determine the fields required for research. The Project Administrator should be able to specify the ResearchSpace default 'house' taxonomies or taxonomies available from other projects. Data automatically converted using the ResearchSpace business process gateway would also ingest the underlying taxonomies using the SKOS standard with the CRM framework.
- 12.3.3 The potential extent of supported fields depends upon the current status of the ResearchSpace data template and what has already been mapped to the CIDOC-CRM. As such, a project will configure the online form by selecting the available fields for presentation to data providers who will use it. The form will provide a facility for creating new records, saving data in the RS staging area and then modifying or deleting it before indicating that the record is ready for import into the system. This should be done by assigning an appropriate project/RS status in the RS Catalogue. New records are highlighted in the Project Administrator's Dashboard and the Project Administrator can monitor the process and see when the status changes to indicate that the data is ready for assigning a project status. Fields not supported by ResearchSpace should be added to the appropriate template and mapped to CIDOC-CRM. The online form should operate with the same web service used for automated upload of data.
- 12.3.4 It should be possible to configure this tool so that entries are alerted to the Project Administrator through the dashboard system and that a link be provided to allow the project lead to view new data entries. The online form should also be available for project members within the ResearchSpace environment.⁷

12.4 Asset Upload

- 12.4.1 As with the online data entry form a separate online form should be available for upload of documents and images. The form should provide both individual upload with appropriate metadata and status fields as well as bulk upload with a simple CSV metadata mechanism which would hold the metadata for multiple assets along with the local location of the file. For these bulk uploads the CSV file would be processed and the location used to upload the associated asset file.

12.5 Terminology

- 12.5.1 The quality of the searching and browsing facilities depend upon how well the information can be harmonised. The CRM provides a primary and fundamental platform for this harmonisation but ResearchSpace also requires that data be harmonised at the level of taxonomy. While data imported into ResearchSpace will share a common structure and vocabulary at a high level it will have different taxonomies applied to it. This approach to this is twofold:
- Use a default or base set of taxonomies for all new data where the submitter has no preferred taxonomy model.

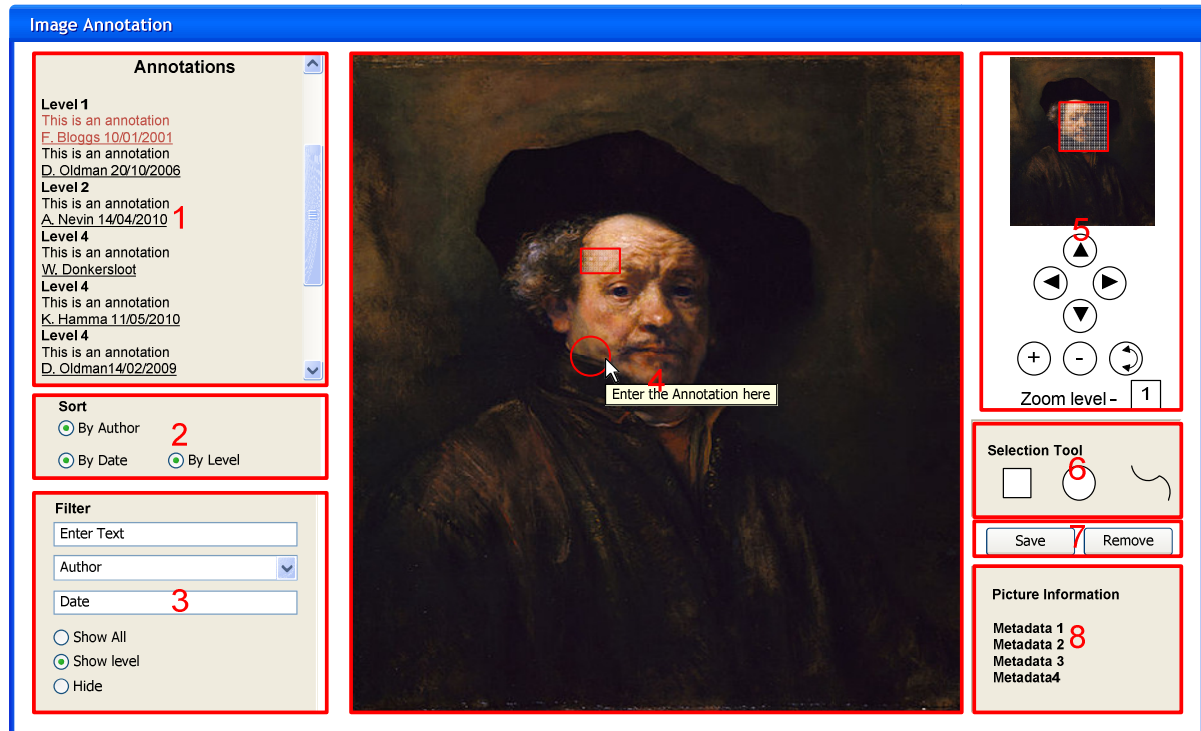
⁷ It should be based on the internal data input and annotation tool described above.

- Provide helper tools to so support the mapping (co-referencing) of different terms.

Annex 1 – Research Tools Functionality

Note: Drawings do not show all functionality

12.6 Image Annotation / Zoom Image Annotation



Area 1 – Annotation View

1. Show previous annotations with date of entry and author. Can be sorted and filtered.
2. Selecting an annotation should highlight the region on the image. Multiple annotations can be selected.
3. Selecting an annotation at a particular level of zoom will cause the image to zoom to that level and the area to which the annotation resides.
4. The user can view an annotation at different zoom levels.
5. It should be possible to show only project annotations.

Area 2 - Sort

1. The list of annotations in area 1 can be sorted by author date and level.

Area 3 – Filter and display options

1. Displayed annotations can be filtered by keyword, author, institution and date range. It should be possible to select multiple authors and institutions.
2. A calendar control should be provided for the start and end date range.
3. Filter should also include annotations at particular levels.
4. Annotation regions should be proportionate depending upon the zoom level so that, for example, a region defined at zoom level 3 will be proportionately smaller at zoom level 1.
5. When the image is zoomed only those annotations defined within the area should be displayed.
6. It should be possible to hide all annotations.
7. Colour coding of annotations to reflect different author's and institutions.

Area 4 – Image Work Area

1. The Image itself can be zoomed and moved using the tools (Area 5).
2. Annotations can be applied using the annotation tools (Area 6). Users can defined a region (rectangle, circle or free form) and write an annotation which then appears in the annotation view (area 1). Annotations should be saved or deleted using command buttons (area 7).
3. Removing the actual annotation from the image means that the image can be view properly without large amounts of text impairing the view. Regions can also be hidden using the hide feature (area 3).
4. It should be possible to enforce ontology terms for the type of annotation. (Open Annotation Project).

Area 5 – Zoom Controls

1. To control the zoom level of the main image.
2. A zoom control is required based on a streaming tiled method (e.g. Zoomify, IIP).
3. Zoom in and out.
4. Pan vertically and horizontally at any zoom level.
5. Should display the current zoom level and the maximum zoom level.

Area 6 – Region Selection Tools

1. Rectangular, circle and custom region selection options. User should select one then drag region onto the image.
2. Tool can be deselected by clicking the region option again.

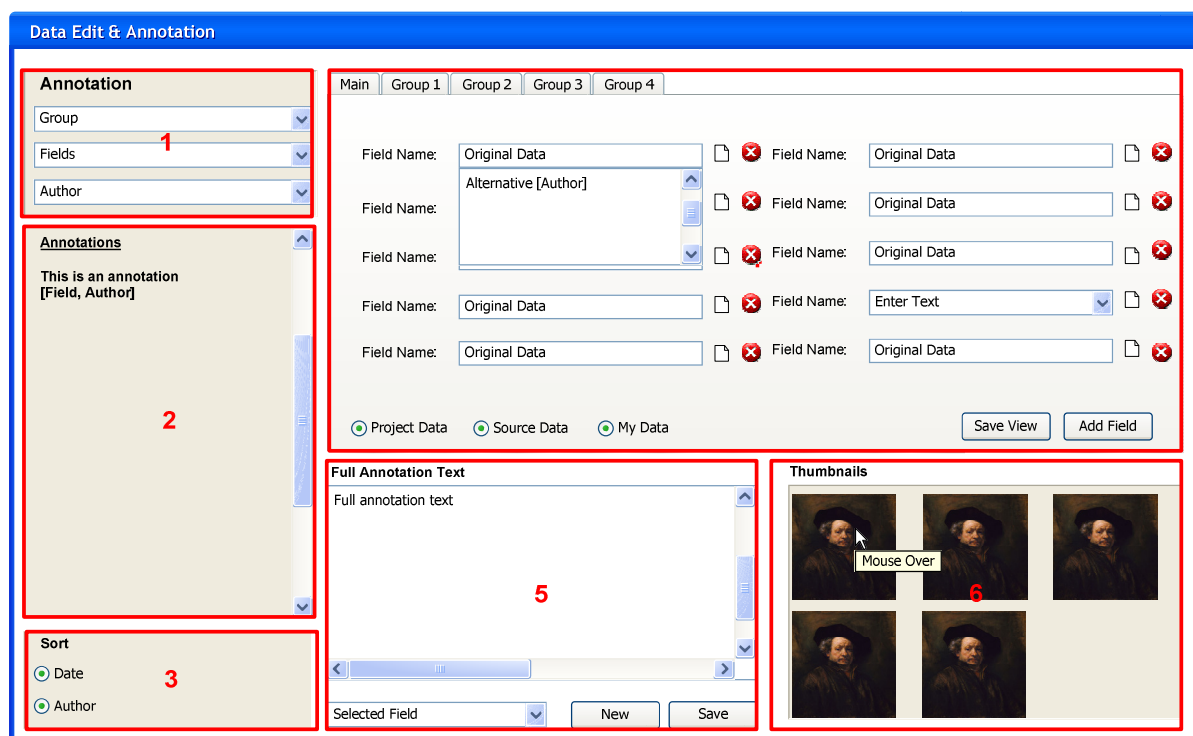
Area 7 – Command Buttons

1. Save – To save the annotation.
2. Delete – To delete a selected annotation (with authorisation).

Area 8 – Image Information

1. Display information about the image loaded into the tool. (Name, Description, image size, format, owner).

12.7 Data Edit & Data Annotation



Area 1 – Annotation Filtering

1. Filter the abbreviated annotations in area 2 by group (of fields), field and author.
2. It should be possible to show only project annotations.

Area 2 - Abbreviated Annotations

1. Shows previous annotations created and referenced against the data fields.

Area 3 – Annotation Sort

1. Sorts the abbreviated annotations by date and author.

Area 4 – Field View and Edit

1. The field view is by group (groups to be determined for each record type). By default the fields should show the original uploaded data (first version). Other versions can be displayed by expanding the combo box which are sorted by date and also provide the author identifier.
2. Each field also has an icon for creating another data version of the data entry. This should open another window which provides access to controlled terms (separate design).
3. The view of fields can be altered to show the preferred project fields, the user's fields and the original source fields.
4. The user's view of the fields can be saved for the next session.
5. The groups will depend upon the type of record selected from the search. These will be different depending upon whether the user is looking at objects, sites or buildings.
6. Button allows deletion of a selected data entry (with authorisation).

7. Can add a new field (Leads to a field definition control for determining the type, size and authority for the field).

Note: Preferred project fields are determined by the Project Administrator through the Dashboard system.

Area 5 – Annotation Detail

1. The text box shows the complete annotation.
2. This can be edited by the original author (and others authorised) and re-saved.
3. Other users can create a new annotation against the selected data field and save that annotation to join the others.
4. It should be possible to enforce ontology terms (qualifiers) for the type of annotation. (Open Annotation Project).
5. Editing should be supported by a full html editing control which includes the ability to insert hypertext links.

Area 6 – Thumbnails

1. Show the thumbnails associated with record.
2. Thumbnails can be used to link to the Image Annotation tool.
3. Thumbnails should provide basic information about the image (file size, type, filename, ID).

12.8 Controlled Data Entry

Area 1 – Field Entry

1. For controlled data entry (would require controlled vocabulary from area 4)

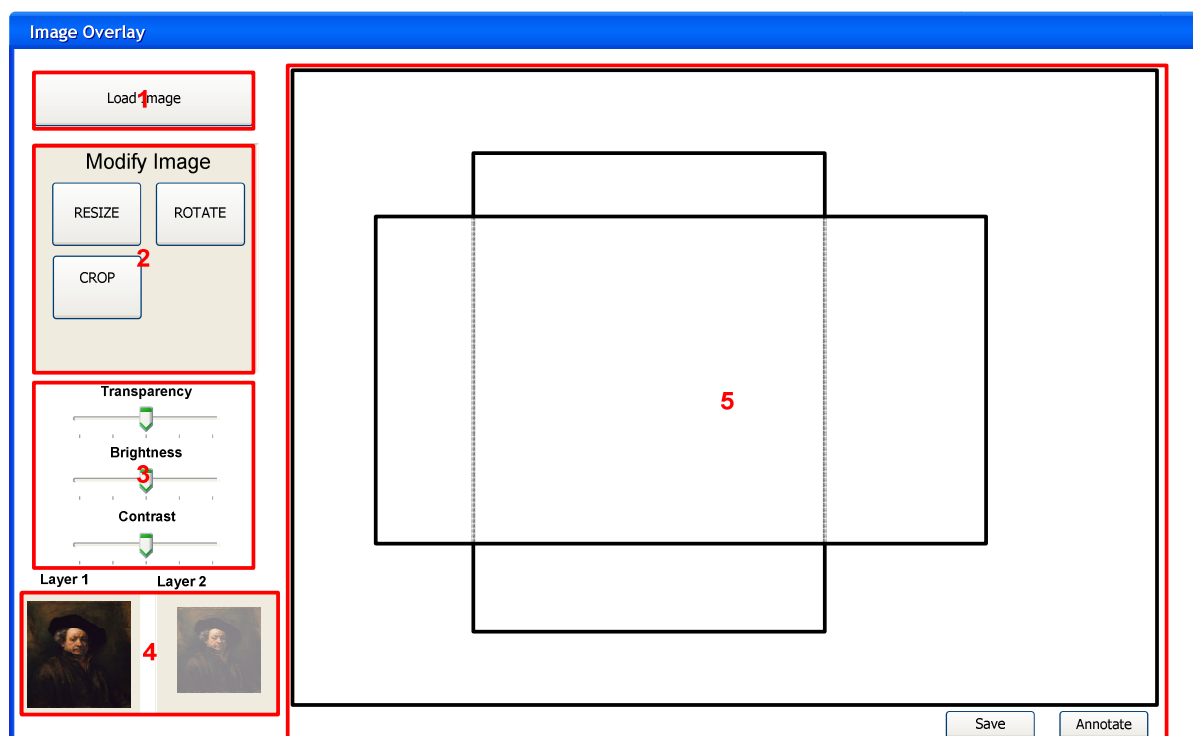
Area 2 – Free text data entry

1. For free text descriptive field entries

Area 3 – Controlled Term Selection

1. Loads associated thesaurus automatically based on data field.
2. Can select authority from drop down.
3. Vocabulary can be flat or hierarchical.
4. Can search vocabulary for a term.

12.9 Image Overlay



Area 1 – Load image

1. The image loader should allow upload of an image from a local computer or allow a user to select an image from a ResearchSpace Digital Asset library.
2. The loader should allow two images which are loaded into two layers.
3. Should be possible to switch the two layers.

Area 2 – Image Modification Tools

1. These tool operate on the selected layer / image (see area 4)
2. Resize allows an image to be resized using image handles.
3. Rotate allows an image to be rotated using a rotation handle.
4. Crop allows an image to be cropped using image handles.
5. Ability to change images to black and white (greyscales).
6. Should provide precise scaling and resizing tools.

Area 3 – Slider Controls

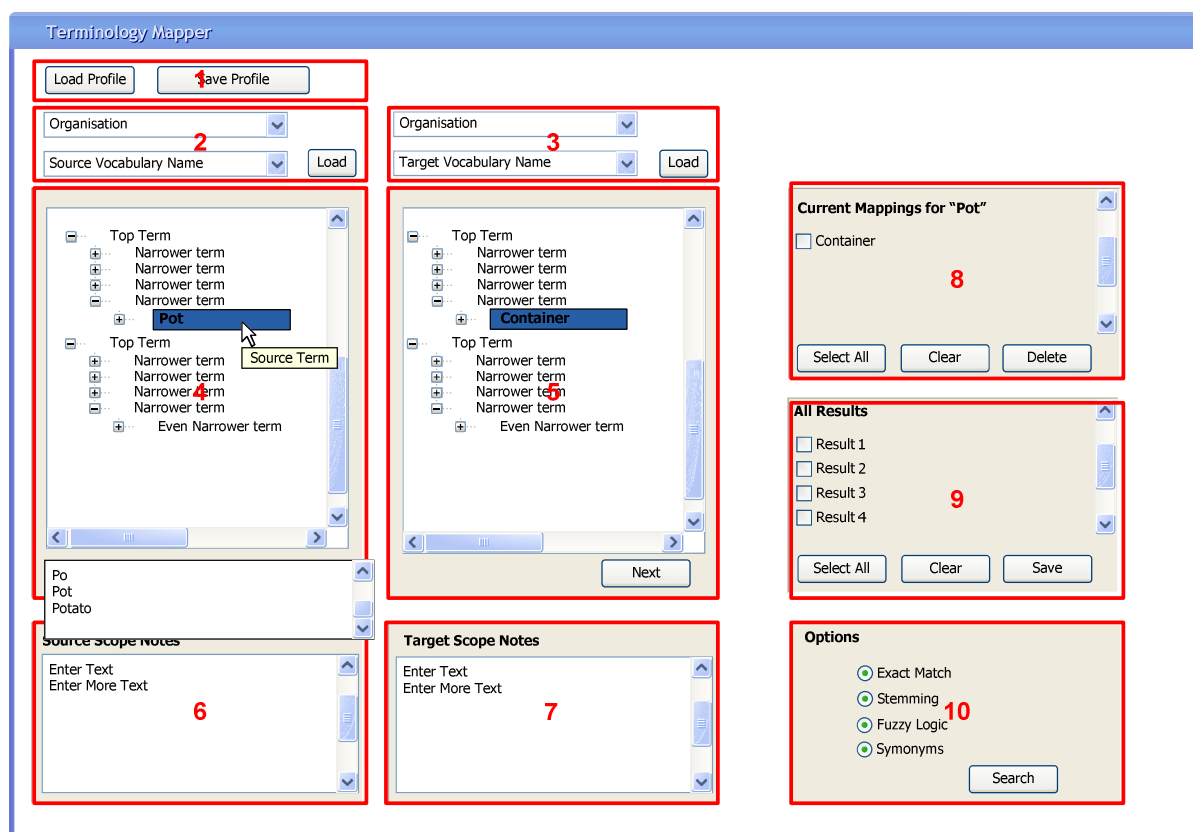
1. Transparency slider alters the transparency of the selected layer / image.
2. Brightness slider alters the brightness of the selected layer / image.
3. Contrast alters the contrast of the selected layer / image.

Area 4 – Layer Selection

1. Images are loaded into layers and each is represented here.
2. The user can select the thumbnail to select the layer that they wish to manipulate. Selecting a layer highlights the thumbnail background to show the selected image.

Area 5 – Image Editing Area
<ol style="list-style-type: none">1. Layers can be moved by dragging a selected image.2. The save button will combine the layers and allow the user to save the image in a ResearchSpace asset library of their choice or save to local disk.

12.10 Terminology Mapper



Area 1 - Load Profile

1. Each project has a profile and each user has an individual profile per project.
2. 'Load Profile' allows a user to load the project profile or their user profile.
3. Profiles should be XML and it should be possible to export mappings to a file.

Area 2 - Source Organisation and Vocabulary Name

1. Choose the organisation from which the thesaurus was sourced. Note: When data is uploaded this may include the authority data files that should be converted into SKOS as part of the conversion. This data will ordinarily be loaded into the shared RDF repository. This project tool will still need to know whether this thesaurus is applicable to the project. This will be the job of the RS Catalogue which needs to record the thesauri owner on either upload or conversion.
2. The name of the thesaurus used as the base vocabulary for searching against another vocabulary.

Area 3 - Target Organisation and Vocabulary Name

1. Choose the organisation from which the target thesaurus is chosen.
2. The name of the thesaurus used as a target for a terminology search based on a source term.

Area 4 - Source Vocabulary

1. This control shows the full source vocabulary either as a flat list or a hierarchical thesaurus.
2. A user can select a term which then becomes the basis for a search.
3. A search mechanism allows the user to find a term and provides alternatives as the user types.
4. The user can manually explore the vocabulary and expand and select using a normal tree control if

the vocabulary is hierarchical.

Area 5 – Target Vocabulary

1. This control shows the full target vocabulary either as a flat list or a hierarchical thesaurus.
2. Terms are selected as a result of a search based on the source vocabulary. Where there is more than one result, multiple terms can be selected. The user can cycle through all these matches using the next button. The user can deselect results if not required.
3. The user can manually select terms (and multiple terms).

Area 6 – Source Vocabulary Scope Notes

1. The text box will show the scope notes for the term in focus on the source vocabulary.

Area 7 – Target Vocabulary Scope Notes

1. The text box will show the scope notes for the term in focus on the target vocabulary.

Area 8 – Current Mappings

1. Shows the mappings already associated with the source term in focus.
2. These terms can be deleted either individually by selecting individual mappings and clicking delete, or by using the 'Select All' button and deleting all mappings. A 'Clear all' button is available for clearing selections.

Area 9 – Proposed Mappings

1. The select results can be mapped to the source term by selecting and clicking the Save button. The mappings are saved to the current profile.
2. Mappings can be selected individually or by using the 'Select All' button.
3. A 'Clear' button will deselect all selected results.

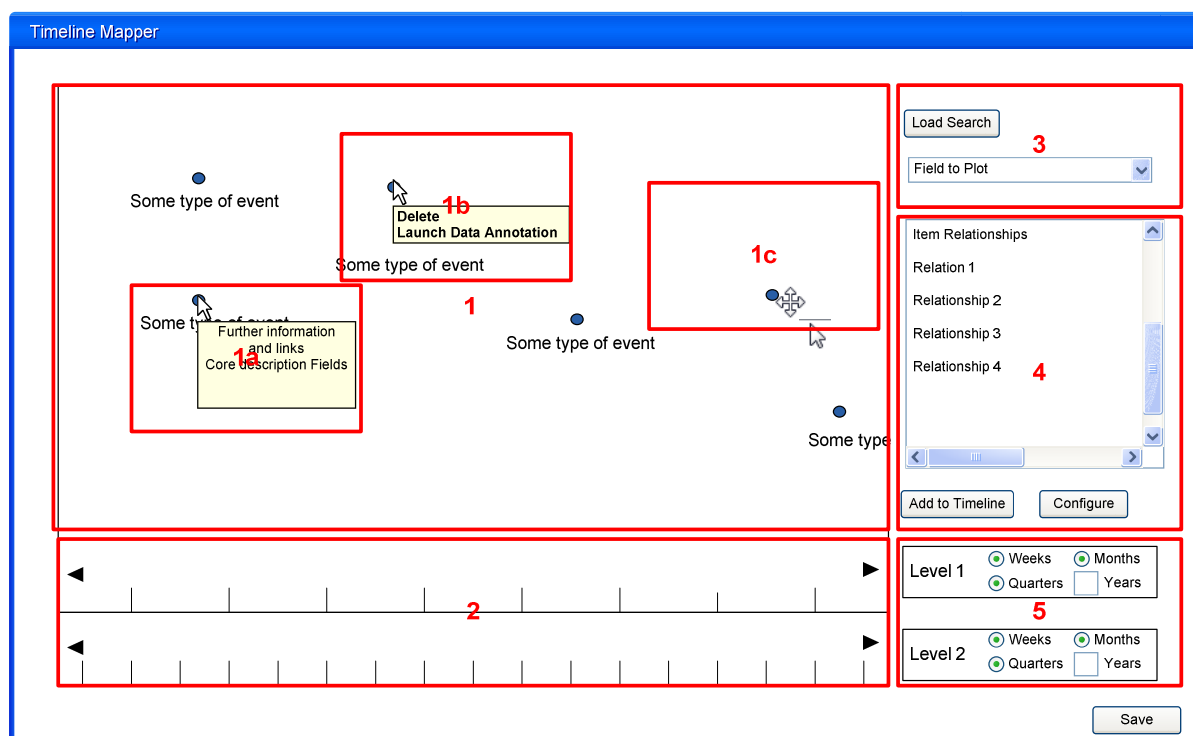
Area 10 – Options

1. The user can select a number of options that will alter the results. These include;
 - a. Exact match – only an exact match will be returned
 - b. Stemming – the term can be match in any part of a target term
 - c. Fuzzy Logic –
 - d. Synonyms –
 - e. Auto match – (see below)

Auto Match

1. The automatically maps of the source terms with target terms according to the settings provided.
2. Mapping profiles should be capable of export to an xml format for editing outside ResearchSpace.

12.11 Timeline Mapper



Area 1 – Plotted Items

1. A results set saved from the Semantic Search tool can be saved and loaded (Area 3) into this tool which will display the items along a timeline defined by 2 levels of timescale (area 2).
2. The plotted items, when clicked will show a box with further information about the item. The box should be closed with another click to a close symbol located on the upper right hand side of the box (1a).
3. A right click should reveal a quick menu allowing an item to be deleted from the display. It should also allow the user to launch the data annotation tool for the selected item (1b).
4. The user should also be able to drag a plotted item to a new date (the date should be displayed as the item is moved so that a precise date can be chosen. This should create a new data in the underlying data under the name of the user as if the data annotation tool was being used.
5. The user can drag the whole timeline along moving it forwards and backwards in time.

Area 2 - Timescales

1. The timeline has two timescales to show two different scales of time that can be configured to suit the data being used. These are configured using the controls in area 5.

Area 3 – Load data / Field parameter

1. The timeline plotter works from a dataset created from a specific search which has been saved in the ResearchSpace semantic search tool. This can be loaded into the plotter.
2. The user can select multiple date fields to be used for the plot.

Area 4 – Relationships

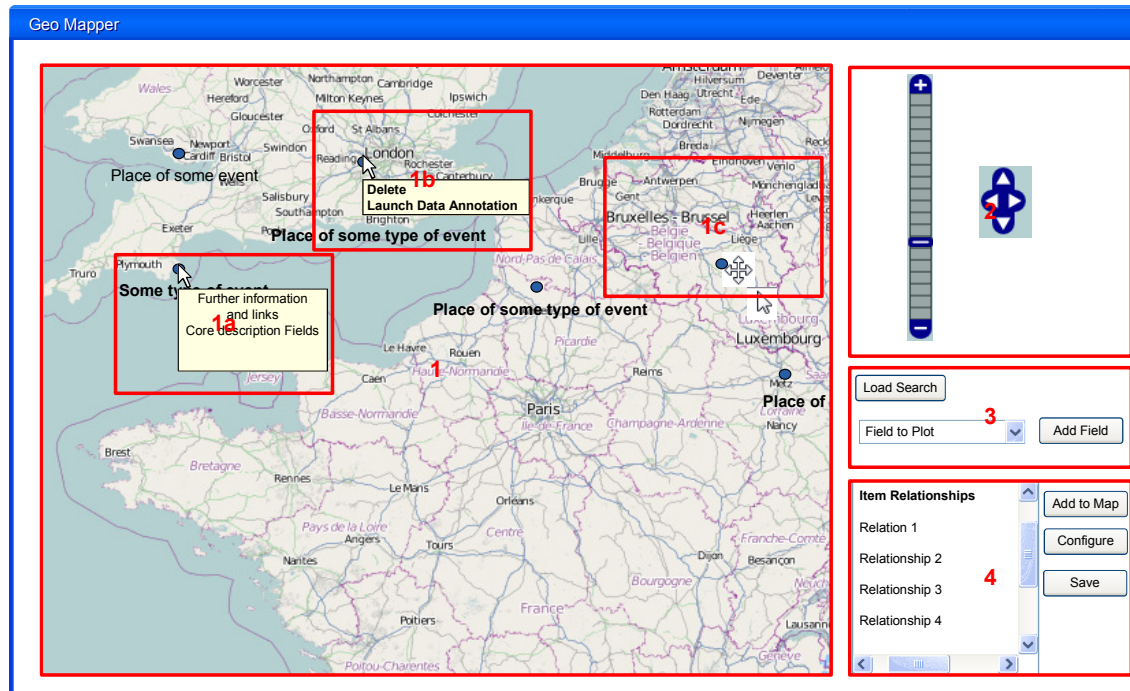
1. When an item on the time line is selected this control will determine related items not in the current data set but which may be of interest. The user should be able to configure this relationship based on particular metadata.

2. These can be added to the time plot by clicking the, "Add to Timeline" button.

Area 5 – Timescale configuration

1. The configuration options allow each timescale to be set with units of days, months, quarters and years.

12.12 Geographical Mapper



Area 1 – Plotted Items

1. A results set saved from the Semantic Search tool can be saved and loaded (Area 3) into this tool which will display the items on a map. The status and configuration of the map (it plots and zoom level) can be saved to an xml file using the save button (area 4).
2. The plotted items will show basic information but when clicked will show a box with further information about the item. The box should be closed with another click to a close symbol located on the upper right hand side of the box (1a).
3. A right click should reveal a quick menu allowing an item to be deleted from the display. It should also allow the user to launch the data annotation tool for the selected item (1b).
4. The user should also be able to drag a plotted item to a new position (the coordinates and place names should be displayed as the item is moved so that a precise location can be chosen. This should create a new data in the underlying RS or project data store under the name of the user as if the data annotation tool was being used.

Area 2 – Map Control

1. The two control allows the map to be magnified and for the user to navigate around the map.
2. Controls should also be available on the mouse and keyboard.

Area 3 – Load data / Field parameter

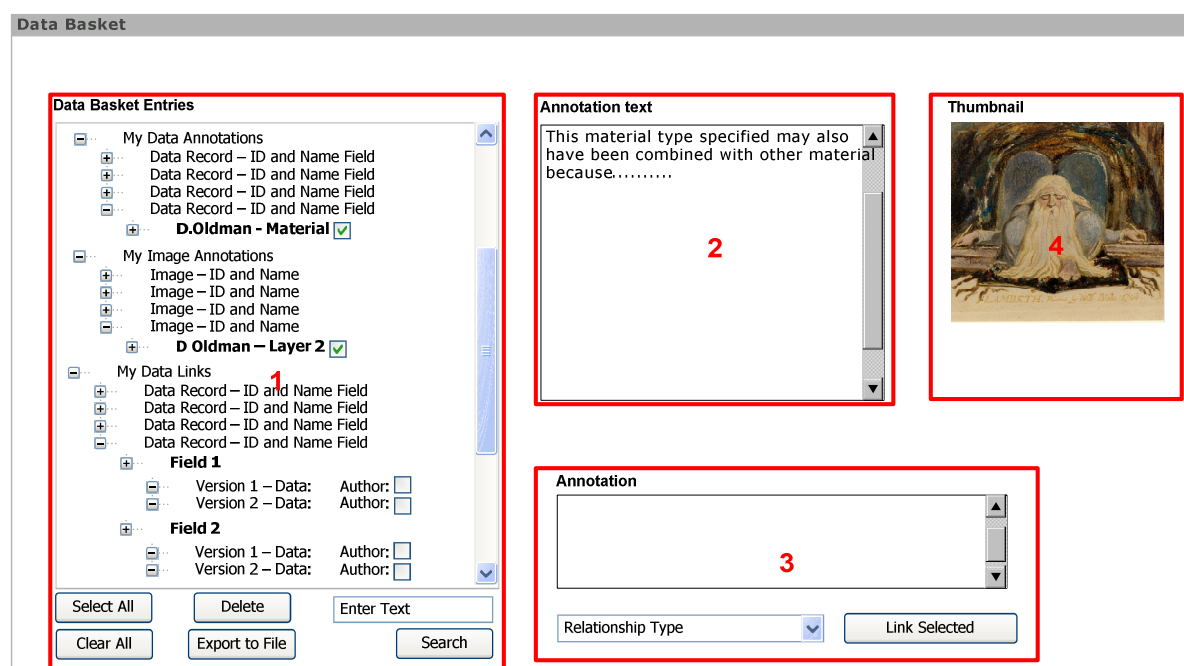
1. The geo plotter works from a dataset created from a specific search which has been saved in the ResearchSpace semantic search tool. This can be loaded into the plotter.
2. The user can select multiple geographical fields for plotting items.

Area 4 – Relationships

1. When an item on the map is selected this control will determine related items not in the current data set but which may be of interest. The user should be able to configure this relationship based on particular metadata.
2. These related items can be added to the map by clicking the, “Add to Map” button.

3. The state of the map can be saved for another session.

12.13 Data Basket



Area 1 – Data Basket Entries

1. Data copied from the Data and Image Annotation tools is reorded here under different sections.
2. Entries can be selected in two ways, by a tick box and by highlighting.
3. Highlighted entries can be deleted and the whole data basket can be cleared.
4. Allows capture of data records, images (through a link), data annotations and image annotations.
5. Enables data basket entries to be selected for insertion into social networking tools. (Insertion into a social networking tool should also include a URL link for another user to view the data using the original research tool).
6. It should be possible to find entries in the data basket using keyword searching, author, record ID and field type.

Area 2 – Annotation and Data text

1. Displays the annotation or data text for a selected item.

Area 3 –Data Relation Link

1. Entries can be semantically linked using an Annotation qualifier determining the relationship. (Open Annotation Project).

Area 4 – Thumbnail

1. Displays the Thumbnail which is the subject of an image annotation or the image associated with the relevant data record.

Annex 2 - Technical Architecture

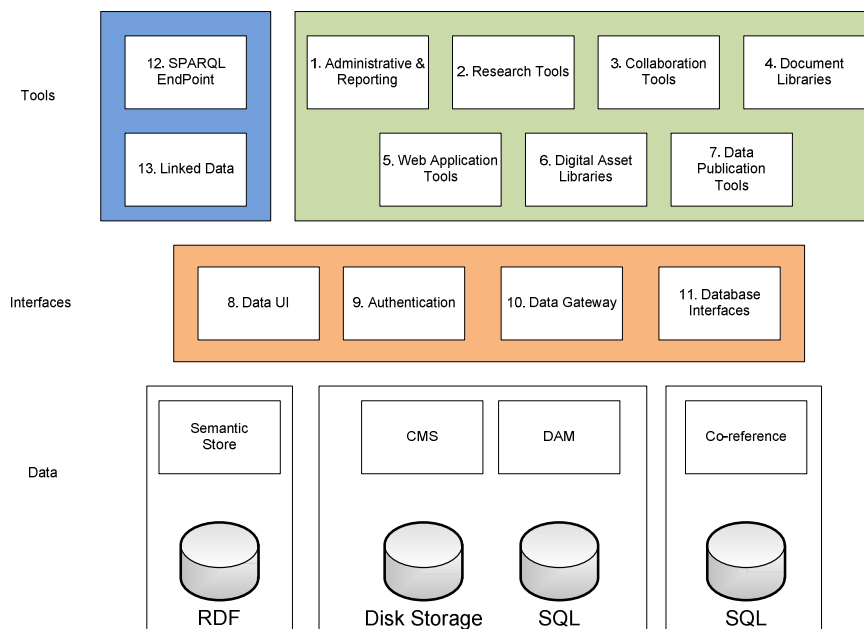
13 Introduction

This annex describes the ResearchSpace architecture explaining how the different architectural components interact and support the functional specification. It is based on the premise that an open source content management system will be used as the base system for the ResearchSpace environment. The short listed CMS evaluated as part of stage 2 of the project are Nuxeo, Alfresco and Drupal and all have some degree of media management. As such statements regarding the architecture take into account the services and architecture provided by these three systems.

13.1 Architectural overview

The key components within ResearchSpace are collaboration, research tools and data, particularly from an RDF store. Collaboration tools are provided as part of the base CMS platform while research tools are new components that must exist and interact within the same environment but must also directly interact with the RDF system as independent tools in their own right. As such, lightweight coupling is preferred; any specification or implementation which requires components to be closely integrated is to be avoided wherever possible. Interfaces between components should be clearly defined, and offer methods that perform simple, concise, and discrete actions, as is typical in a service-oriented architecture.

The following high-level diagram shows a simplified version of the layered architecture proposed for ResearchSpace. Briefly, there are three conceptual layers – at the lowest level, data storage and management facilitates; at the uppermost level, the applications, tools and outwardly facing components, whether they are hosted as part of ResearchSpace or external tools/visualisations; and in between an interface layer which provides core functionalities and controls and mediates access to data, storage and digital assets within the ResearchSpace environment.



13.2 General principles

As an open system, closed protocols, proprietary APIs and formats are to be avoided. Where not specifically stated in this document, the following rules should be applied. Any exceptions to the following must have a clear business case, such as that an existing component would require too much resource to comply.

The ResearchSpace environment will primarily be accessed through the a web browser interface, however specialised research tools which are not web-based may additionally be adapted to access data from and contribute to ResearchSpace as required.

- All data, components and services must be capable of being invoked over HTTP.
- The primary mode of access is by URI resolution to a simple data URL (document) or Linked Data URI (semantic data).
- Services are invoked wherever possible in a REST-compliant manner; if full REST compliance is not possible, then REST-like URIs must be used.
- Complex queries in this REST-like style using standard SPARQL will be supported.
- All metadata shall be represented in an accepted format of RDF, such as RDF/XML, Turtle or N3, in order of preference.
- For data objects, such as documents and images, open formats such as html, jpeg and mp3 are preferred, although it is accepted that proprietary formats such as PDF and Microsoft Word may sometimes be necessary.
- It may be permissible to invoke a service on the same directly, for example by using direct invocation via functions, but only if there exists an equivalent means via HTTP; the direct route being used solely for performance or similar reason.

13.3 ResearchSpace Components

The ResearchSpace components considered by this annex are;

1. LDAP Provider – Authentication and Authorisation
2. Shared and Project Specific Stores
3. ResearchSpace Registry – The central repository of Project information
4. ResearchSpace Catalogue – The central catalogue of research data
5. Document and Asset Store – Libraries of documents and multimedia files
6. Social Networking Tools – Main user area
7. Research Tools – Tools to operate over data and assets
8. Text Searching
9. Co-referencing services

14 ResearchSpace Authentication and Authorisation

14.1 Introduction

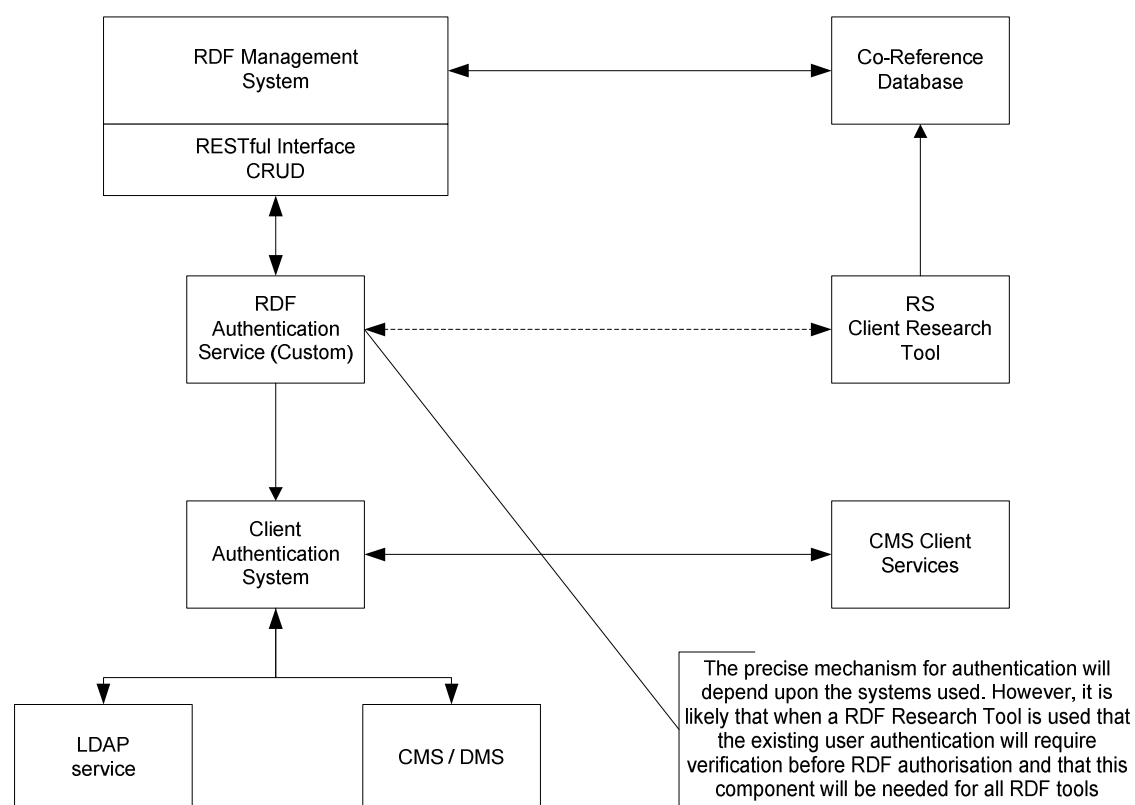
The ResearchSpace system is an online environment accessed using web browser. Authentication needs to initiate the authorisation of the components provided by the Content Management System, including the CMS social networking tools, document libraries and asset libraries, as well as authorisation to the RDF project store to which the project has access mainly through the research tools. All projects are authorised to use the shared CMS and RDF components and data stores. All authentication and authorisation should occur through single sign on.

14.2 Purpose

- To enable users to login and connect to the correct project resources. When users logon they should arrive at their Project dashboard from which they can access other resources and find out about project activity.
- To ensure seamless interoperability between the CMS and RDF components and allow single sign on between the components.

14.3 Scope

The linked data public interface will be available for all to query (the shared ResearchSpace repository) without authentication but under an agreed licensing policy.



14.4 Architectural Process

Research tools use a separate RDF database which is independent from the CMS database and system. However, a user of ResearchSpace will need to authenticate against both the CMS system and the RDF database. The preferred method for providing authentication across both systems is LDAP. The preferred CMS systems for use with ResearchSpace include Alfresco, Nuxeo and Drupal. Both these systems can be integrated with a LDAP service. LDAP can provide the basis for single sign across both systems but the mechanism by which this will provide authentication for the independent RDF database system depends on the final RDF system selected for ResearchSpace. However, it is likely to be one of the following;

1. The RDF store is exposed as a web service and uses web server authentication. In many cases the web server can be interfaced with the same LDAP service used for the CMS system. (e.g. 4Store)
2. The RDF store has its own internal user authentication system. If there is no direct LDAP interface then one would need to be developed (e.g. Sesame).
3. The RDF application has an API that allows interface with LDAP (e.g. Virtuoso).

The architecture will be as follows;

- The user logs on once using the CMS logon mechanism.
- The logon is directed to the LDAP authentication system.
- The authentication returns a user ticket indicating the groups that the user belongs to.
- The user requests services from the CMS which are authorised by the LDAP service
- The user requests services from the RDF Store which are authorised by the LDAP service

14.5 Interfaces

- CMS Login Point – The physical login environment is provided by the CMS.
- CMS Project Environment – A successful user logon allows the user to progress to the project environment created within the CMS.
- Project Collaboration Tools – CMS uses LDAP to authorise internal security for the social networking tools.
- Project Document and Asset Libraries – CMS uses LDAP groups to provide authorisation to the shared and project document libraries.
- Shared Document and Asset Libraries – CMS uses LDAP groups to provide authorisation to the shared and project asset libraries.
- Project RDF Store – RDF management system uses LDAP to determine access to project stores.
- Shared RDF Store - RDF management system uses LDAP to determine access to the shared store.

The functional requirements incorporate the ability to transfer data between the CMS environment and the research tools. The requirement is that users should be able to embed and create links to data (originating from the RDF store), documents and assets.

14.6 Additional Controls

Although a user may have access to one or more project stores and therefore one or more collaboration areas, the user must only be able to access one project at a time. If a user is a member of more than one project they will be asked to choose their context for their session.

14.7 Security

The ResearchSpace system authenticates using the Login mechanism provided by the Content Management System (CMS). When project areas are created the appropriate groups will manage the 3 main roles for project

access. The appropriate CMS security will be mapped to these groups for the CMS components that the users interact with. RDF store security has a simple read/write type access model.

SSL encryption should be configured for login and LDAP communication (LDAPS)

14.8 Question and Answer

How do users authenticate in ResearchSpace?

Users authenticate using a Content Management System login plug-in. This should be configured to use LDAP rather than the CMS SQL database?

How are usernames and passwords protected?

The login process should use SSL encryption (https) as should an LDAP connection (LDAPS).

How do users access CMS facilities used in ResearchSpace such as collaboration tools (forum, blog, wiki) and document and asset libraries?

Once users are authenticated they would use the inbuilt security system for reading, writing, etc. The environment would be setup to deploy these rights through CMS roles, ResearchSpace Administrator, Project Lead and Project Member. These are the default roles for ResearchSpace.

How will external research tools work with ResearchSpace?

Tools created independently of ResearchSpace can operate against the linked data end point provided. In order to work against the raw RDF new research tools must be incorporated into the ResearchSpace environment and its authentication system. Because RS tools are independent of the CMS and its API it should take relatively little effort to incorporate an RDF standards based tool into the system.

15 Shared and Specific Stores

15.1 Introduction

ResearchSpace has the concept of both shared and project repositories. The shared repository is open to all projects and RS project members. Linked data from the shared repository is also available to the public through a linked data interface and SPARQL endpoint. Documents and Assets within the shared repository will only be available to the visitors through web applications that make them public ally accessible, again through an agreed licence. Project data and information will only be available to project members until it is transferred to the shared repositories.

15.2 Purpose

The purpose of the repository system is to ensure that projects can complete their research agenda and publish their findings without having to make the data and information they generate available before that point. Keeping project datasets in separate named graphs allows permissions to be enforced so that only project members can access a project RDF repository. Once the project agenda has been completed then this work should be made available to other projects working in the field, whether working within a ResearchSpace project or externally.

15.3 Scope

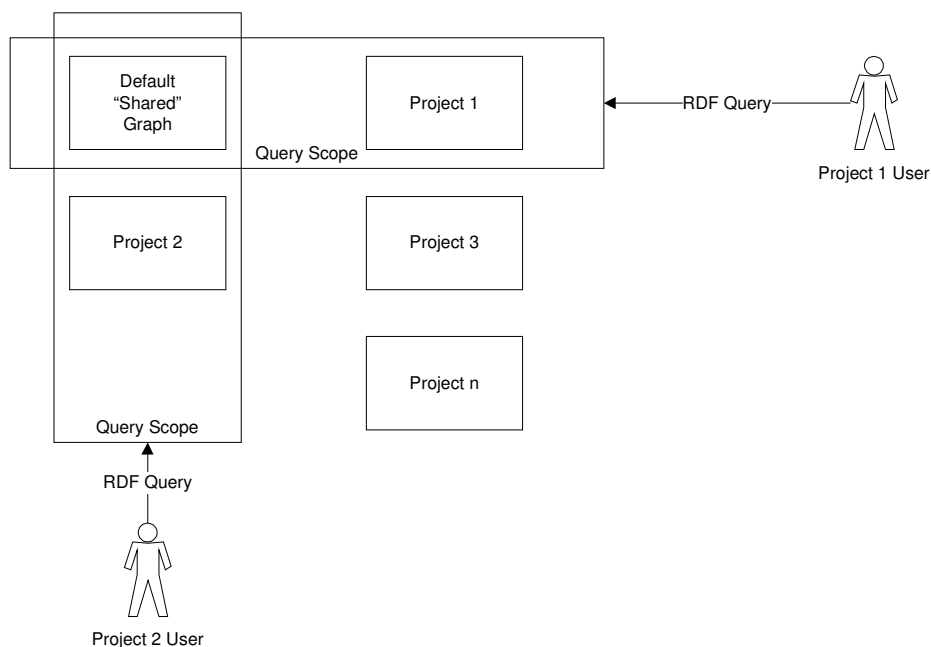
- External project will only have access to a shared repository linked data interface configured appropriately.
- Projects must be able to query RDF data across both their own project RDF store and the shared RDF repository. Similarly, project should be able to search for documents and digital assets across project and shared repositories.
- Projects will not have access to other projects data, document or digital assets.

15.4 Architectural Process

RDF data will be maintained in named graphs. Each project will have its own named graph alongside a named graph for the shared RDF repository. It is anticipated that ResearchSpace will operate one endpoint service for all named graphs and this means that queries can be federated across 2

named graphs (an RDF dataset comprising of the project named graph and the shared named graph) through one service. The default named graph is the shared repository.

The new SPARQL 1.1 standard provides more options for federated search techniques and these should be considered before final implementation.



15.5 Interfaces

The ResearchSpace Registry contains a list of all the named graphs in the repository.

15.6 Additional Controls

N/A

15.7 Security

Security to named graph will depend upon the system employed but should be at the endpoint and graph level.

16 ResearchSpace Registry

16.1 Introduction

The RS Registry is the administrative database for ResearchSpace and records information about the Projects.

16.2 Purpose

The RS Registry allows those governing ResearchSpace to correctly administer projects and to allow others to know what projects are being carried out on ResearchSpace to encourage further collaboration between projects whether carried at on ResearchSpace or elsewhere.

16.3 Scope

The RS Registry provides all information to ResearchSpace administrators, project information to project members and more limited information to the public.

16.4 Architectural Process

Initial information is completed on a project request to use ResearchSpace and can be expanded during the project lifecycle.

16.5 Interfaces

ResearchSpace and Project Dashboards and a public information web page.

16.6 Additional Controls

N/A

16.7 Security

Different Registry information will be available to the different ResearchSpace roles.

17 ResearchSpace Catalogue

17.1 Introduction

The RS Catalogue keeps a record of data movements into and within the ResearchSpace environment.

17.2 Purpose

To record the inputs and outputs of data supplied into the different data and file stores. The RS Catalogue effectively provides an inventory of data and information held by each project and within the shared repository.

Data and information can be uploaded, converted and modified into ResearchSpace at which point the process should be recorded with the appropriate status. The RS Catalogue should record the person who has uploaded the information (the ID of the person should be linked to the RS Registry where the person's organisation data will be stored), a description of the data or file, the repository that it has been loaded into, the size of the data load and the status of that data.

17.3 Scope

- Recording newly created or uploaded asset files (documents, digital assets).
- Recording the movement of these assets between repositories.
- Recording uploads of data to ResearchSpace (using metadata supplied on upload)
- Recording movements of data from the Project repository to the shared repository (using metadata mandated by the system to describe what is being moved).
- Record conversions to structured data undertaken within ResearchSpace (using metadata mandated by the user interface).
- Record depreciated triples and any metadata associated with that decision including user information and reason for depreciation.

Version information on document assets should be recorded by the inbuilt features of the CMS.

17.4 Architectural Process

The RS Catalogue needs to be part of all processes involving data movement and these processes will require UI features that ensures that the necessary information is included for a true audit record. This includes the Data Gateway component (upload and conversion) and logs of movements of data between different document and asset libraries. Projects should be able to generate a report listing the project resources and their locations including assets provided by the project but stored in shared repositories. It is possible that this data is an extraction of data already stored by the CMS.

17.4.1 Uploaded data

When users upload data through the ResearchSpace gateway they should be required to enter metadata to describe the data they are uploading, the ownership details and the status of it. The upload process should either allow this information to be inserted into an associated html form or should be embedded into the data file being upload (which should be validated before upload). This header information includes the repository that the data is being uploaded to. By default it will be the shared repository. However, the user can choose a project repository of which that user is a member and has authorisation to upload data.

17.4.2 Internally Converted Data

When unstructured data is uploaded to ResearchSpace the project members may attempt a conversion of data within the ResearchSpace environment itself. In this event the RS Catalogue should record the source of the conversion.

17.4.3 Moving Data

Data can be moved when it leaves a project store and enters a shared store. It can also be converted from files stored in document libraries into data within the RDF store. These actions should also be stored and where appropriate linked to previous catalogue entries that have some relationship with the moved or converted data.

For example;

- A document is transferred to a different document library.
- A document is converted into RDF from a file stored in ResearchSpace.
- Data is moved from one named graph to another

Document version control is a function of the CMS system.

18 Document and Digital Assets Store

18.1 Introduction

The content management system used for ResearchSpace should be capable of supporting document and digital asset libraries. The metadata associated with these files will automatically be stored in the Content Management System but should also be stored as triples within the triple store so that it can be associated with related data. The triple store should also reference the http image resource address.

18.2 Purpose

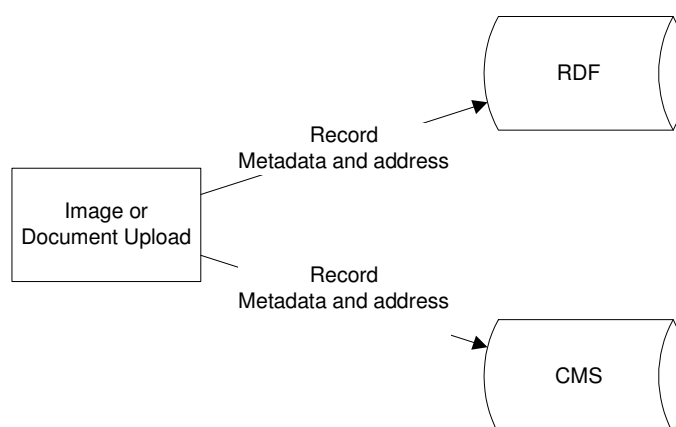
The purpose of document and digital assets stores is to provide ways of storing these files in logical containers for ease of use. These will be of particular use to projects which may wish to organise project documents and assets according to a particular structure

18.3 Architectural Process

18.3.1 Projects

When documents or images are uploaded to ResearchSpace, metadata will be mandated to describe the files. As a minimum this must include the file name, a description, Keywords to be associated with the file and the copyright information. It is expected that these mandatory fields can be configured within the CMS. The resource information should also be recorded in the RDF system. Project administrators should also be able to configure additional metadata fields on a mandatory or optional basis for their project.

Version control should also be configurable through the standard features of the CMS.



18.4 Interfaces

- ResearchSpace Catalogue records project resources and their movement.
- CMS and RDF databases record metadata about the resource.

18.5 Additional Controls

N/A

18.6 Question and Answer

What happens when documents or images are transferred to shared libraries?

When assets are transferred to a shared repository from a project repository the libraries will be subject specific rather than project and subject specific. This may be achieved using metadata and content type information. Therefore the appropriate shared library should be found (with consultation with ResearchSpace administrators) to locate the correct stores to copy images. Metadata regarding copyright, the project that they came from and other descriptive metadata and tags should be preserved.

19 ResearchSpace Toolbar

19.1 Introduction

In order to aid bringing the various components of the ResearchSpace environment together to form a cohesive user experience, a common *ResearchSpace Toolbar* shall feature throughout all web-based collaborative and research tools. The toolbar shall facilitate common and easy access across the environment to user dashboards, present notifications and alerts, and enable both collaborative and research tools to be easily launched.

The Toolbar shall be implemented by means of adding a JavaScript 'include' to each web-based tool, which shall in turn dynamically request, configure and inject the toolbar as appropriate into HTML pages. The scripts to generate the toolbar shall be centrally maintained at a predetermined location on the researchspace.org domain.

19.2 Interfaces

The toolbar should be available within collaboration and research tools.

20 Co-reference Services

20.1 Introduction

Each project may have one or more sameAs services, representing equivalences between different mappings or expressing different contexts. Individual project members could also have additional personal sameAs stores; however this is likely to be unnecessary for the majority of users and only used in advanced scenarios. On the whole it is anticipated that a more usual workflow is to permit project members to contribute to a single notion of equivalence which is shared across the project.

A sameAs service is agnostic to the types of the URIs it manages – it can be used to represent other predicates such as owl:equivalentProperty in the same way as owl:sameAs.

In the context of ResearchSpace, the sameAs store is used for managing the following mappings:

- Between different thesauri;
- Within a single thesaurus;
- Linking NIR (Non-Information Resource) equivalence between datasets;
- Linking NIR equivalence within datasets;
- Ontology equivalence to non-RS ontologies.

20.2 Component Location

The sameAs services are sited entirely on the main RS server, accessed via Data Interface.

20.3 Interactions within the ResearchSpace environment

Research tools and other components will make frequent lookup requests to sameAs stores via the Data Interface, as well as facilitating the management and addition of data to sameAs stores.

The Data Interface requires access to Authentication Interface and ResearchSpace Registry to determine which sameAs stores to access.

20.4 Inputs / triggers

The ResearchSpace Data Interface shall provide a REST-like service such that given a specific resource or concept, additional equivalent concepts may be retrieved. The Data Interface shall function as a meta-service, aggregating results from a number of sameAs stores, as determined by user preferences and permissions accessible via the presence of an authentication token.

Components within the ResearchSpace environment may request that data is added or removed to a specific sameAs store, given appropriate authentication. URI equivalences are created by pair-wise assertion, with an optional accompanying comment. This is the sole way of asserting equivalence knowledge into the sameAs store. The sameAs service can roll-back assertions to the point at which an erroneous pair was asserted

ResearchSpace Administrators and Project Leads may, via their dashboards, request administrative operations on a sameAs store, including setting accessibility options and the set-up/deletion of an entire store.

20.5 Outputs / actions

The outputs of a sameAs query are a set of URIs, including the submitted query term. The query can be parameterised (either in the URL or by Content Negotiation) to request the format for the returned bundle, such as RDF/XML, text/N3, and results returned accordingly.

The sameAs service is passive, and does not initiate any actions on other components.

21 Social Networking Tools – Main user area

21.1 Introduction

It is currently envisaged that ResearchSpace will use social networking tools already available with the selected content management system. The social networking tools are the hub of the ResearchSpace environment forming the basis of collaborative discussion with links out to documents, assets and data. This means that the architecture must facilitate interfaces with the different data and administrative components. Some of these may already be inherent within existing CMS functionality.

21.2 Component Location

Social networking tools are applications within the CMS system.

21.3 Interfaces

1. The ResearchSpace toolbar - providing navigation application launch capability
2. The RS Data Basket – a mechanism for managing data to and from the social networking tools.
3. The Research Tools – which are launched from the CMS environment and which interface with the CMS environment through the data basket.
4. Navigation to other ResearchSpace components.
5. Reporting engines for reporting activity.
6. Search Engine (searching within active social networks)
7. Through hyperlinks (used to link and locate data, documents and assets and as parameters for launching Research tools).
8. Dashboard

22 Research Tools – Tools to operate over data and assets

22.1 Semantic Search and Explore

The Data Interface shall, as described earlier, provide a single point of access to enable research tools to easily query all applicable semantic data. This will have the appearance and behaviour of a standard SPARQL 1.1 Endpoint, and by default return queries against only public data sources. If a query is made and supplied with a ResearchSpace authentication token, for example via a research tool being used by a logged-in project member, then the query should be executed against all data resources that the individual users is permitted to access.

The exact implementation of this functionality is yet to be determined, as it is heavily dependent on the capabilities of the chosen underlying semantic data store. However, there are two approaches which potential solutions could broadly take. In either case, *named graphs* shall be used to enable metadata to be recorded against sets of triple statements, detailing for example when that set of triples was asserted, by whom, from what source, and to which project they are associated. The first storage scenario is that private data is held in completely separate repositories, one for each project or source. The query endpoint in the Data Interface shall then have to query across the applicable repositories and aggregate results. The second storage scenario is that all data is held in a single repository, and subsequently query execution patterns and results are inspected to ensure that they do not use triples from named graphs for which the user does not have permission to access. Neither of these functionalities are likely to be natively supported by existing triplestore systems, hence the requirement of a “semantic library” to implement the required access controls.

The “separate store” methodology potentially offers a more scalable and generic solution, as data can be spread across different repositories, on different servers and even utilising different triplestore implementations, which can be swapped in and out as desired. In essence, middleware “glue” sits above a number of generic triplestores to perform the cross-repository querying. However the down-side is that there are possible constraints on the expressivity of queries that can be executed across different repositories, depending on the complexity of the middleware library developed to achieve this. In addition, there are likely to be trade-offs between the opposing desires of performance and capability.

Conversely the “single store” approach is likely to be less restrictive in the expressivity of queries that can be executed, as it is essentially a native store, and this would remove the performance issues in cross-repository querying, presuming that all data can itself be efficiently loaded into a single store. However to implement the required named graph checking is likely to be a complex task involving non-trivial modifications deep within the triplestore code base. This will require both an intimate understanding of the chosen triplestore implementation, and result in a solution which is non-transferable and could potentially have a significant performance impact itself if not carried out extremely carefully with due regard to the inherent optimisations within the existing code.

22.2 Component Location

The Data Interface is a key component within the ResearchSpace middleware layer. Services shall be accessible from researchspace.org server, however triplestores are likely to be hosted on a separate dedicated server(s) within the ResearchSpace environment.

22.3 Interfaces

- Receive queries from ResearchSpace tools, components and external sources.
- Handle query execution in the manner required.

22.4 Inputs / triggers

- Receipt of queries to SPARQL endpoint

22.5 Outputs / actions

- Queries are on the whole passive operations, however see next section regarding data update.

23 Semantic Update

While most SPARQL queries are read-only, some triplestore implementations support special query verbs which can upload, modify and delete data, however available functionalities vary. If these are to be supported, in order to permit research tools to make arbitrary adjustments to the semantic data stores, then such modification events must be noted and recorded appropriately in both the named graphs storing provenance, and the ResearchSpace Registry and Catalogue as required. In addition, new data must be passed to the text search indexing processes.

The more controlled approach to regulating data upload and modification is to wrap these functionalities in simple services exposed through the Data Interface. For example, annotation tools and the collaborative utilities within ResearchSpace have the common requirement to assert that an annotation or discussion post has been made within the environment. These share common requirements to recording the provenance and detail of the event, and as such can easily be encapsulated in a service permitting them to make such assertions by means of a simple request without the tools having to make any semantic manipulations or understanding whatsoever.

Such services implemented within the Data Interface for these purposes shall accept the necessary parameters encapsulating the details required, and then execute the appropriate assertions, modifications or removals into the requisite triplestore, catalogue, event log and text indexing service. It is recommended that this is the best practise for facilitating data acquisition from tools which otherwise do not require the handling of semantic data, and of course removes the necessity for each tool to then call the various sub-services such as text indexing or maintaining the RS Catalogue. Tools and services that already deal with semantic data can also utilise similarly defined services to request the assertion of a chunk of RDF data. The Data Interface should additionally intercept SPARQL queries containing data modification verbs, and handle these updates, as appropriate, before passing them directly to the underlying stores, as required.

In add cases, the presence of a user authentication token is required to check that sufficient permissions are granted, and to provide the necessary additional context provenance information.

23.1 Component Location

The Data Interface is a key component within the ResearchSpace middleware layer. Services shall be accessible from researchspace.org server, however triplestores are likely to be hosted on a separate dedicated server(s) within the ResearchSpace environment.

24 Relationship Mapper

User interaction to establish and maintain correspondences between terms.

The user has a context in which they are working, which will provide default values for many of the configuration parameters, and may also restrict the range of optional values for these parameters that is offered.

This is a component that allows users to explore one or more EP/NGs, to construct a view of equivalence between resources in those EP/NGs. For RS, it is specialised towards thesaurus entities (which will be represented using SKOS), but it should be built in such a way that it can be used for resources of any type, by specialisation if necessary.

The component interacts (if necessary) with the user to identify the EP/NGs to be used.

It interacts (if necessary) to identify which thesauri are to be considered.

It interacts to identify which Mapping is being maintained or created, and if it is being maintained, it is loaded into the component to inform the matching process.

The matching interaction is then an on-going process.

At any time, and at the end, the component may decide to save the results of the interaction to the sameAs store. It is desirable that the tool asserts each match as it is made – this will allow the user to switch to a different tool and see the effects of their work so far.

Access to each EP/NG is governed by the role and access rights of the user to those components.

Access to the sameAs is similarly governed by the role of the user.

The computational costs and even network demands of this component are small, and so the location of the code logic can depend on the technology being used – for example a web application generator will often provide a standard deployment pattern.

24.1 Component Location

Server code serving web app, with logic located between browser script and server-side script as appropriate. Browser preferred.

24.2 Interfaces

- Data Interface (SPARQL, sameAs, data assertion)
- Query RS EP/NG for a list of EP/NGs (*admin*) – [optional auth.].
The main RS EP/NG maintains a record of valid EP/NGs, from which the tool can set the contexts.
- Query RS EP/NG for a list of EP/NGs (*admin*) – [optional auth.].
The main RS EP/NG maintains a record of valid EP/NGs, from which the tool can set the contexts.
- Query EP/NG1 for thesaurus terms (*RS*) – [optional auth.].
Using the selected thesaurus, the tool needs the terms so that it can provide a view of them for the user to choose candidates to match against.
- Query EP/NG2 for possible matches (*RS*) – [optional auth.].
As the user selects each candidate, the tool may find possible matches from the target thesaurus by finding existing semantic relationships that are relevant..

- Query TS for possible matches (Text to *RS*) – [optional auth.].
By querying the text search service using any labels from the candidate term, possible matches can be presented to the user.
- Query EP/NG1 and EP/NG2 for labels and contexts (*RS*) – [optional auth.].
The tool needs to give the user a visualisation of the context of the terms being matched. This may well depend on the types of terms being matched, and may require queries and resolution to other targets.
- Request new sameAs from the Admin service – [auth.].
Starting on a brand new mapping.
- Query EP/NG1 for sameAs and mapping list (*admin*) – [optional auth.].
Each EP/NG has a record of the mappings that are valid for that context; this is to select the sameAs store and mapping that is being worked on by the user at the moment.
- Query sameAs for matches (*sameAs*) – [optional auth.].
The tool needs to give the user a view of the existing matches in that mapping.
- Post output to sameAs (*RS + sameAs*) – [auth.].
One or more terms may be posted to sameAs, along with any notes on the match, and metadata about the match (user, tool, date).
- Post retract to sameAs (*RS + sameAs*) – [auth.].
Previously asserted pairs may need to be retracted as views change or mistakes are identified.
- Authenticate with auth. service.
The tool established the authentication rights, and at any stage must cope with an access failure because of authentication failure of a particular access.

24.3 Inputs / triggers

User interaction via web browser.

24.4 Outputs / actions

- Assertions to sameAs store
- Request to Data Interface to record user actions

25 ResearchSpace Text Searching Component

25.1 Introduction

Although RDF provides a flexible data model and allows both powerful searching and browsing, it is not optimised for searching literals. ResearchSpace will require a search engine to ensure that search performance is optimal and to provide the following:

- Aggregation functionality
- Filtering features.

25.2 Interfaces / Resources Affected

1. The shared ResearchSpace RDF repository
2. The project RDF repository
3. The Content Management System – (Social networking data)
4. The Content Management System – Document & Asset Libraries

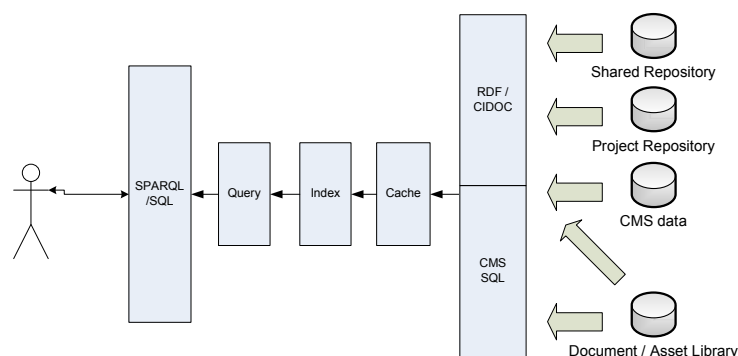
25.3 Implementation Options

As a result of the market evaluation of open source CMS the following options are available.

1. RDF databases, CMS database and documents are all indexed using a single search index engine, most likely Lucene/Solr. (Preferred option)
2. Indexing is performed through a combination of Content Management System document indexing facilities and the use of Lucene/Solr.
3. Documents are indexed through the CMS engine but metadata about the documents is indexed through Lucene/Solr

The advantage of option 1 is the potential to introduce interference to the index and make connections between documents and databases within the indexing system. See

<http://fgiasson.com/blog/index.php/2009/04/29/rdf-aggregates-and-full-text-search-on-steroids-with-solr/print/>



25.4 Purpose

The text searching facility will be used for the following activities;

1. To improve the performance of searching triples that have free text strings in both shared and project repositories. This may include authorised data fields as well as annotations that are linked to a record ID.
2. To provide text searching facilities across ResearchSpace documents and abstracts for which the user has access.
3. To allow users to search through text written within the ResearchSpace collaboration / social networking tools.

25.5 Number of Indexes

1. Document Index – The CMS should index documents on entry to the native document library
2. Shared ResearchSpace Index
3. Project repository indexes

25.6 Usage

1. Semantic Search Tool (data and images) – to identify data and images.
 - a. The user can elect to perform a simple uncontrolled text search.
 - b. The user can search text areas using a keyword.
 - c. The user can search documents
2. Relationship Mapper
 - a. The user can search SKO triples and scope notes
3. Document Search Tool
 - a. The user can search within documents as well as descriptive summaries within the CMS provided when the document was uploaded to a document library.

25.7 Index Update

1. On document upload – Documents should be indexed on upload into a ResearchSpace document library.
2. On document metadata update – When metadata about a document is saved
3. On image metadata update - When metadata about a digital asset is saved
4. On conversion to CIDOC CRM – When new data enters one of the RDF repositories.
5. On new Annotation – When RDF annotations are saved.
6. On new data version – When new record metadata is saved.

25.8 Access

Text searching is limited by user security access and to the repositories that are available to the user. A project team member can search their own project store and the shared research project store. The results must indicate which store the data refers to.

A system account should be used to add new data to the indexes.

Configuration

- RDF / CIDOC schema fields requiring indexing
- Document Metadata fields requiring indexing

Use of Fuzzy Logic, Stemming and Synonyms

[To be completed]

Annex 3 – Component List

<p>Online Project Request</p>		<p>Create Project</p>	<p>ResearchSpace RD Store</p>	<p>The Online Project Request provides initial information about a new project filled out by the requesting project. The Administrative part of the form only available to ResearchSpace Administrators is completed internally.</p> <p>Data entry should be on one form.</p> <p>Data Inputs / Outputs</p> <ol style="list-style-type: none"> 1. Project Name – Each Project should have a unique RS identifier* 2. Project URI – If the project has an Internet domain name 3. Brief description of the Project and Project approach 4. Lead Organisation(s) – There could be one or more lead organisations per project* 5. Partner Organisation(s) – There could be one or more partner organisations actively involved in the project 6. Other organisation(s) – There could be other contributing organisations not actively involved in the project. 7. (Address details for all organisations) 8. The Principal Investigator(s)* 9. Principal Investigator(s) Email Address* 10. Principal Investigator(s) Telephone* 11. Project Contract name 12. Contact Email Address 13. Contact Telephone 14. Business Address Details 15. The Funding Bodies(s) – could be finance by partners themselves 16. One contact detail for each funding body 17. Indicate which organisations are proposing to upload data 18. Estimate extent of multimedia assets 19. Overall storage estimate 20. Estimated Project Closure Date <p>ResearchSpace Administration Section</p> <ol style="list-style-type: none"> 21. Decision of ResearchSpace Board (Accept, Reject, Pending) 22. Decision Notes 23. Project Status (ongoing) 24. Status Notes 25. Project Subscription Amount <p>Other Behaviour</p> <ul style="list-style-type: none"> • Ability to update data subsequently to initial submission • The form should trigger an email to ResearchSpace Administrators. • The request should appear on the RS Administrator Web Dashboard as a new request with a drill down to the detail.
--------------------------------------	--	-----------------------	-----------------------------------	--

Administrators Dashboard	Create Project / Collaboration / Close Project			Description <ul style="list-style-type: none"> • Welcome Page and Project Information: • The latest activity and of changes • Access to RS Registry • Provide a list of projects (Summary and Detail) • Edit mode to change field values • User Accounts • Full account administration • Access to RS Catalogue • Data entering and leaving the ResearchSpace (Project and Period) • Quotas • Capacity Statistics • Messaging functionality • The ability to view and send messages to ResearchSpace users. • Alerts and Audit Information
Project Dashboard	Project Collaboration			Description <ul style="list-style-type: none"> • Welcome page and Project Information • Latest announcements and news for the project. • Latest activity • Users activity • Links to previous activity area. • Project Communication • Create project news and announcements • Access to RS Registry • Provide a list of projects (Summary) • Drill down to see information for his/her projects • Provide an edit mode to change field values • User Accounts • Full administration of Project user accounts • Access to RS Catalogue • Data entering and leaving the project data stores • Status of project data and information • Capacity Statistics • Configuration • Access policy to the site • User setup and notification • Social Networking component setup • Workflow configuration • Document management library setup and configuration • Asset Management configuration • Creating the research questions and objectives • The ability to move assets from Project stores to ResearchSpace stores • List of entry annotations made on the project with links to them • List of documents stored in the system by the

				<p>user</p> <ul style="list-style-type: none"> List of images stored in the system by the user Messages to and from other ResearchSpace colleagues Ability to determine whether new metadata fields can be created by research team Alerts to the dashboard for new metadata fields <p>Views / Screen Reporting</p> <ul style="list-style-type: none"> Activity by User Account – posts, data entries and annotations made by each project member Activity by User Account – Detail. Showing all the entries made Links back to the entry in context. A filter should also include publication status. Activity per component or discussion thread. A filter view of all the inputs per collaboration tool. <p>Messaging</p> <ul style="list-style-type: none"> Ability to initiate a message to a user (or users) from a data entry such that it is included in the message. (e.g. when looking through a particular view the project administrator selects a particular input and initiates a message screen which automatically incorporates the entry and a link to the entry) <p>Workflow</p> <ul style="list-style-type: none"> ResearchSpace tools that write data including annotations should be referenced in the dashboard tool such that all new entries can be reviewed by the project lead(s) and categorised appropriately. The activity reports should provide the ability to indicate the status of new data and annotations. <i>[Would it be sensible for this to be administrative data pointing to RDF rather than in the RDF itself?]</i> <p>System</p> <ul style="list-style-type: none"> Storage Used and Left Data Uploads Completed No of Images uploaded by organisation <p>Control</p> <p>The project administrator should be able to control the status and access to particular collaboration areas. For example, it should be possible to close a particular collaboration.</p> <p>The project administrator will have access to all</p>
--	--	--	--	---

				<p>areas, but can also delegate access to different dashboard areas to other project team members on a read and write basis.</p> <p>Behaviour</p> <p>The dashboard will provide a menu or tab for each functional area.</p>
User Dashboard				As above with some limitations
Online Data Entry Tool		Upload		<ul style="list-style-type: none"> • Upload source data • Upload documents and assets with appropriate meta data • Wizard for dynamic field selection and form generation • Data validation • Ability to extend field support • Mapped to CIDOC-CRM • Dashboard alerts and links • RS Catalogue status flag • Ability to review, edit • Fields not supported by ResearchSpace should be added to the appropriate template and mapped to CIDOC-CRM. • Support for house or organisation terminology vocabularies.
ResearchSpace Toolbar		Collaboration	Configuration data store	<ul style="list-style-type: none"> • Available throughout the ResearchSpace environment. • Added to all collaboration editors • Launch available/ configured research tools • Access project document and asset libraries • Access their dashboards • This should be done without losing the state of the collaboration editor
Navigation Environment		Collaboration	CMS / SQL	<ul style="list-style-type: none"> • Provide a consistent and focused mechanism from moving between the main ResearchSpace elements • Tight navigation between main ResearchSpace elements • Provide intuitive links between the administrative components of ResearchSpace (from the Dashboard) and the data in context • Ensure that data and asset activity leads back to the collaborative environments • The navigation system should provide connections between the different key components. The collaboration environment should provide access to the research tools from the data posting screens and the navigation must lead the user back to the same point in the collaboration tool • Users should be able to easily switch between related tools. For example if annotating

				images it should be easy to switch to data annotation to the record(s) to which that annotation refers.
Data and Link Basket				<ul style="list-style-type: none"> • Tool used for transferring information between Research tools and collaboration areas (two way) • Links and data can be brought back from relevant research tools • Annotations can be brought from collaboration areas to relevant research tools • Data and links imported from research tools should be identifiable and reportable from collaboration areas.(should be able to easily identify what data has been brought in.)
Research Tools (General Note)				Users should be able to launch research tools and open data records or images that they have used in the past (recent history).
Semantic Search Tool				<ul style="list-style-type: none"> • Ontology Searching – with the labels used on the CIDOC-CRM ontology framework. • Thesauri Searching – with thesauri terms to search for data records • Option for using different thesauri and mapping profiles. • Faceted search using thesauri • Free text search • The searching system should incorporate auto-complete, alternatives, synonyms, stemming and other search techniques. • Results • Summary results should be a thumbnail and the following fields [to be completed] • Results should include the remaining information about an object organised according to the original source data and then the annotation and versions of data below. • The tool should indicate what field annotations or new versions refer to so that these are not easily duplicated. • From the search tool the relevant research tools should be available for launch • Users should be able to mark records for other research tools. For example, create a basket of data records for timeline or geographical mapping. • Relationships to both ResearchSpace and external data sources.
Data Annotation Tool				Data Input tool <ul style="list-style-type: none"> • Record display form with editing features • Create new version of data • Create Annotations for existing data

				<ul style="list-style-type: none"> • Attribution and audit information. W3C standards • Appropriate predicates for type of annotation • Review and correct converted data before generally available • Navigate through all the records • Configure the field's displayed and switchable to full view.
Relationship Mapper				<ul style="list-style-type: none"> • Compare and map terminology between flat or hierarchical authority files • Configurable matching techniques including: <ul style="list-style-type: none"> ○ Word Stemming ○ Synonyms ○ Fuzzy Matching • Matched both terms and their scope notes • The user should be able to select a term from the results to display the structure (broader and narrower terms) • A term could be mapped against many other terms depending upon how the two thesauri compare in detail and precision. • User can correct or accept mapping and then save the record • Mappings should be displayed. • Different profiles should be possible which may have different mappings for the same terms. • It should be possible to apply different profiles when searching.
Data Gateway Service				<p>Description</p> <p>The data gateway is the mechanism by which data is entered into the ResearchSpace system. Contributing organisations and individuals may have data in the following formats;</p> <ul style="list-style-type: none"> • Organised in an information system database • In a structured file format • In an unstructured file format <ul style="list-style-type: none"> • Templates to which source data can be mapped. • Templates comply with CIDOC-CRM. • Web Service for organisations to convert data into ResearchSpace. • Upload of data documents • Upload of unstructured documents • Tracking within RS Registry • Status <ul style="list-style-type: none"> ○ Public ○ Release on Project End ○ Release during project on authority

				<ul style="list-style-type: none"> ○ Project Only (not for release) ○ Source for RDF data conversion only (not for general research use) ○ Converted ○ Project Approved Document ○ Approve Converted document – status for PI to approve. ○ Superseded by (another document) ○ Withdrawn ○ Deprecated • Other meta data <ul style="list-style-type: none"> ○ Copyright status ○ License ○ Person who loaded ○ Organisation • Previous annotations of deprecated data survive new RDF data assertions and can be retained or deleted. • Audit/ version indicates when users make changes • RS Registry for processing status, indicating changes (differences) • (audit and version – service by CMS but data reflected in RDF) • Important status changes may need approval, and are configurable by Project Administrator. • RS Catalogue need to ensure that there is a reference between source and converted documents and vice versa
RS Registry				<p>The database that stores information about the projects which is initiated from the online request form.</p> <p>Available from Dashboards.</p> <ul style="list-style-type: none"> • Decision of ResearchSpace Board (Accept, Reject, Pending) • Decision Notes • Project Status (ongoing) • Status Notes • Project Subscription Amount • The ResearchSpace Registry should be available for public viewing although it should be possible to configure which Registry fields are included.
RS Catalogue				<p>A record of data and assets contained within ResearchSpace together with the location and status of that data.</p> <ul style="list-style-type: none"> • Record creation, modification, movement, deletion. • Data or Asset Identifier recorded with operator’s user name, and timestamp.

				<ul style="list-style-type: none"> • Dashboard access to logs and configurable dashboard alerts. • Quota and capacity information • Ability to change status from Dashboard • Reports on Registry activity by period, user, data type and status.
Image Zoom (and annotation)				<ul style="list-style-type: none"> • Optional protection against download for high resolution images (decided by Project Administrator) • Displayed next to the image • Show region of the image that the annotation refers to. • Annotate at different levels of zoom. Clicking Annotation moves to the correct zoom level • Zoom into the image to the required degree and embed the annotation at that level. • Annotations displayed for all levels and at current level of zoom (can be toggled) • Provenance shown
Image Comparison				<ul style="list-style-type: none"> • Side by side comparison • Scaling and overlays • Ability to annotate with reference to images within the overlay. • Image adjustment techniques to highlight/show differences between overlaid images.
Linking Tool				<ul style="list-style-type: none"> • A hypertext link can be made from an object record or image to an external web site page. • A hypertext link as an annotation to a piece of data to an external web site page. • An object to object RDF association within ResearchSpace. • An object RDF association to a URI outside ResearchSpace. • Links should also be supported by the ability to add further contextual annotations.
Version Comparison (Track Changes)				<p>ResearchSpace has the opportunity of bringing together a whole body of work on a particular subject. As time goes by more information will be added and in RDF data is always added but not taken away.</p>
Geographical and Timeline Mapping				<ul style="list-style-type: none"> • tool for converting geographical names into coordinates (generally supporting longitude and latitude coordinates information) • User defined data set which would have both time/location data (e.g. Exhibition history (time/location), discovery time/location, purchase history, etc.)

				<ul style="list-style-type: none">• Annotated data set with explanations• Represent time/location and time-frame or geographical areas of interest which would incorporate selected data
--	--	--	--	---

Annex 4 - Reusable Open Source

IIP Image - High Resolution Zoom Software. <http://iipimage.sourceforge.net/>

TimePlot Widget - DHTML-based AJAX widget for plotting time series and overlay time-based events over them. <http://www.simile-widgets.org/timeplot/>

Timeline Widget - <http://www.simile-widgets.org/timeline/>

Open Annotation Project - <http://www.openannotation.org/>

OpenZoom SDK – Software Kit for high resolution image zoom - <http://www.openzoom.org/>

Open Street Map - <http://www.openstreetmap.org/>

Treeview Control - <http://www.jstree.com/>