

Assessment of the ResearchSpace Framework

Functionality and Necessary Adaptations for the Production of
AnthroGraph



Stefan Schlager, Felix Engel

17th August 2018

Stefan.Schlager@anthropologie.uni-freiburg.de

Contents

1 About this Document	4
1.1 Aim and Scope of this Document	4
1.2 Function Within the "e-Research Technologies" Funding Proposal	5
2 Changes To the Code Structure	6
2.1 Creation Of a multilayer Architecture	6
3 Change of Data Model	7
3.1 Retention of Features Specific to ResearchSpace	7
3.1.1 Reconfiguration of Faceted Search	7
3.1.2 Assets Management	8
3.1.2.1 Documents (and other file types)	8
3.1.2.2 Images	8
3.1.3 Clipboard	8
3.1.4 Annotations	9
3.1.5 Semantic Narratives	9
4 Page Creation	10
5 Access Rights	11
5.1 Roles and Users	11
5.2 SPARQL Mode and LDP Mode	11
5.3 Visibility Settings	12
6 Conclusion	13
A metaphacts License and Support Overview	14

Nomenclature

AR	Anthropological Researcher
DFG	Deutsche Forschungsgemeinschaft (German Research Foundation)
HTML	HyperText Markup Language
IRI	Internationalised Resource Identifier
LDP	Linked Data Platform
MVP	Minimum Viable Product
RDF	Research Description Framework
SDR	Researcher charged with Software Development
SPARQL	SPARQL Protocol and RDF Query Language

1 About this Document

This document is about the development of a software which here is referred to by the working title "AnthroGraph". AnthroGraph is a server application that will help scientific institutions to set up RDF-based information systems for managing research data emanating from investigations of skeletal human remains. It is designed as a tool that will help institutions and researchers to integrate semantic research data modelling in their regular research routines.

Within the funding scheme "e-Research Technologies" issued by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), AnthroGraph is to be realised as a minimum viable product (MVP) by adaptation of the ResearchSpace framework. This document assesses necessary changes for this project.

1.1 Aim and Scope of this Document

This document lists functionalities of ResearchSpace that will be employed in the development of the MVP as well as necessary configurations and adaptations. In addition, it also hints at options for further adaptation at later project stages. The document was compiled for the funding application to the DFG and represents the state of knowledge about ResearchSpace at that time. Additional issues, clarifications and corrections are likely to occur during software design, projected for phase 2 of the proposed work programme.

This document does not touch upon the creation of page templates and how they determine navigation through the application. Developing these is the objective of project phase 2. Here, only provision of technological basics is discussed.

1.2 Function Within the "e-Research Technologies" Funding Proposal

This document estimates the amount of effort needed to render ResearchSpace applicable in the context of the funding proposal within the "e-Research Technologies" funding scheme and the scope of tasks that will be required in this process. This information is necessary to set up the work programme and specify qualifications of employed researchers.

The issues discussed in this document are based on the scope document related to the MVP of AnthroGraph that is also attached to the funding proposal.

2 Changes To the Code Structure

2.1 Creation Of a multilayer Architecture

Deploying existing software often makes it necessary to alter and adapt the code. To make these changes transparent and to maintain the possibility to integrate new features and bug fixes from the upstream software development, we favour a multilayer architecture. This means that the original folder structure is mimicked in an own out of source folder, introducing an additional layer containing custom changes and additional functionality only needed by the AnthroGraph system. This could be employed successfully in the previous project where such a solution already existed¹. When deploying the software, the installer script will favour the content of those custom folders over the ones provided by the stock version. The ResearchSpace software does currently not support this option. In the course of this project, we will implement such a multilayer architecture to facilitate the adaptation of the ResearchSpace functionality to our specific needs. The commercial support (see Appendix A) applied for in the funding proposal will guide the Software Development Researcher (SDR) through the process of adapting the installation routine to enable a multilayer architecture.

¹<https://wiki.duraspace.org/display/VTDA/Building+VIVO+in+3+tiers>

3 Change of Data Model

ResearchSpace mainly relies on the classes contained in the CIDOC CRM ontology, supplemented by the classes from the underlying metaphactory framework. In order to provide the adequate functionality for recording anthropological research, the underlying data model needs to be adapted to these ends. In the previous project, we developed RDFBones, an RDF ontology representing key concepts of research in biological anthropology. The Anthropological Researcher (AR) will evaluate, which parts of the ResearchSpace ontology will be needed for a fully functional MVP and which parts can be discarded for the benefit of a slim and efficient data model. As the template pages (see 4) in ResearchSpace are tailored to present the classes of the currently implemented ontology, a major part of the SDR's work will be to create templates based on RDFBones concepts, drawing on the aid of the commercial support service to be acquired.

3.1 Retention of Features Specific to ResearchSpace

ResearchSpace is currently sporting a variety of custom features to accommodate the needs of presenting museum objects. Some of the features (e.g. the faceted search widget) will very likely be useful assets for our MVP, it has to be evaluated by the SDR and the AR, if the adaptation is feasible in the allotted time frame of this project. Other features (e.g. the custom assertions system) seems to add unnecessary complexity to modelling the already complex process of osteological/anthropological research processes.

3.1.1 Reconfiguration of Faceted Search

As mentioned above, the faceted search widget is a very convenient feature for finding contents within the knowledge graph. The SDR, in collaboration with the AR, will be tasked with

remapping the necessary classes to the RDFBones ontology to avoid an entire rewrite of this feature.

3.1.2 Assets Management

Researchspace has, to date, only limited functionality regarding uploading files. There is currently only an upload page for images. But the underlying mechanism can be easily adapted to create upload pages for all sorts of files. The SDR and the AR will have to evaluate how to integrate those filetypes (e.g. documents, 3D imaging data or programming code used for research) into the existing knowledge graph. The following paragraphs outline documents to be supported by the MVP.

3.1.2.1 Documents (and other file types)

While there is currently no specific feature included to deal with text documents, this document type is to be expected as a regular asset of research data. A set of metadata, specified by RDFBones, will need to be implemented for uploads of respective files. This can also work as a generic metadata set for other types of documents.

3.1.2.2 Images

ResearchSpace has an application page for images that allows to upload and manage digital images by adding metadata and specifying semantic relations with database objects. The AR will evaluate if additional features specific to anthropological research contexts will be needed.

3.1.3 Clipboard

The clipboard is a handy feature to keep track of specific objects (instances of classes) by organising bookmarks in hierarchical folder structures. AnthroGraph MVP will retain this feature as it may help structuring the day to day workflows.

3.1.4 Annotations

ResearchSpace has a special feature for annotations. It allows for all sorts of comments on and annotations of existing objects in the database. As this feature is entirely based on using the LDP container approach (cf. section 5.2), its benefits for AnthroGraph will have to be evaluated in the context of the different use case scenarios. If the effort for adapting and maintaining this feature is not feasible, the annotations defined in the RDFBones ontology will be implemented.

3.1.5 Semantic Narratives

Semantic Narratives are a feature in ResearchSpace allowing researchers to write an essay and link it to objects in the database. The use cases defined for AnthroGraph have no scenario where this feature is needed, so it will not be implemented in the AnthroGraph MVP.

4 Page Creation

ResearchSpace allows the creation of application and template pages in a declarative fashion. Template pages define the layout and content for the presentation of instances of specific *classes*, while application pages accumulate content from various knowledge graphs without referring to one single RDF entity only. Their main difference lies in the context of their usage: A template page defines how to present instances of a specific class and technically always refers dynamically to the selected instance. This means that it is only available if the knowledge graph contains at least one instance of the corresponding class. Technically, template pages have a specific variables `''[[this]]''` within the page, and `''?''` in the SPARQL queries that refer to the the IRI of the current RDF resource. The application page, on the other hand, is independent of the knowledge graph's content and can be used to provide overview of an arbitrary assortment of instances. The home page of ResearchSpace, for example, is such an application page.

The templating mechanism relies on HTML5 standards for the client-side components. For security-critical parts of the page, there exists a templating back end based on `handlebars.java`, accompanied by custom helper functions. This allows for distributing the computation between the client for general page generation and the server in order to protect security relevant parameters.

5 Access Rights

5.1 Roles and Users

ResearchSpace relies on the Apache Shiro security framework. This allows to define users, roles and permissions. The platform currently only provides a fine-granular permissions for the SPARQL endpoint, RDFGraph Store, account service and the template service. These permissions are whitelist based, i.e. they have to explicitly be assigned to specific users.

Sets of permissions can be attributed to roles. This allows for an easy per-user rights management by simply assigning predefined roles.

There are currently no options to define groups of users that only have access to specific pages, instead the rights attributed via their roles grant permissions globally throughout the entire platform.

5.2 SPARQL Mode and LDP Mode

ResearchSpace currently stores knowledge graphs (set of triples) in two ways:

1. by adding the new knowledge graph to the globally available graph via SPARQL and
2. storing the new graph in separate LDP containers

While the first possibility has no way of restricting access to sub-graphs for different user groups, the LDP containers would allow for a more fine-grained access management. The downside, however, is that it cannot be used to update existing data or data across graphs (i.e. data not being created in LDP mode).

With the help of the software support applied for in the funding proposal, the SDR will be tasked with evaluation of the applicability of both methods for the defined use cases

5.3 Visibility Settings

Researchspace has implemented visibility settings for template pages of some classes only. This allows to toggle the visibility of the webpages for instances of some classes such as images or annotations. It has, however, no impact on the accessibility of the instance via the SPARQL interface. While it is viable to use this feature for a small set of classes, it is by no means usable to control access to larger parts of a knowledge graph. This, however, relies on using LDP containers (5.2), which comes with the restrictions mentioned above.

For the GUI, however, this feature allows to manage user rights by hiding or showing edit/view buttons on template pages based on the user's properties.

6 Conclusion

While some uncertainties regarding the rights management in ResearchSpace remain, their direct impact on the AnthroGraph MVP is marginal. It is very likely that they can be clarified and solved with the professional aid provided by the commercial support agreement. We conclude that ResearchSpace, and specifically the underlying metaphactory framework, proves to be highly efficient, customizable and provides ample possibilities to solve all kinds of occurring problems. It has to be stressed that most problems regarding the implementation of specific workflows and input/output forms can be handled in a declarative way by using the template pages and without meddling with the source code of the underlying framework(s).

Appendix A **metaphacts License and Support Overview**



metaphacts License and Support Overview

Licensing

Open Source Edition (LGPL) / ResearchSpace

The metaphacts base platform with all its components is generally licensed under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 2.1 of the License, or (at your option) any later version.

The metaphacts platform is bundled with the Blazegraph graph database licensed under the GPLv2 open source license.

Commercial Edition

In addition, we offer a commercial edition that has the following differences to the open source edition:

- Offered with a support option for production use
- Additional features for enterprise use, such as enterprise user management (LDAP) and security

Overview of Editions and Licenses

	Open Source Edition (ResearchSpace)	Commercial Edition
License	LGPL	Commercial License
Graph Database*	Blazegraph	Blazegraph
Available support	Community, developer	Community, developer, production
License subscription fee	free	15.000 € p.a. license subscription (per server)

* Information about support for other graph databases available on request.

Please feel free to contact us in case you need information about other licensing options.



metaphacts

Support

metaphacts offers support on three different levels.

	Community	Developer	Production
Mailing List	✓	✓	✓
Issue Tracker	✓	✓	✓
Direct eMail	×	✓	✓
Private Chat, Telephone call	×	✓	✓
SLAs*	×	✓	✓
Maintenance/ Update support	×	✓	✓
Emergency Patches	×	×	✓
Fee	free	10.000 € p.a.	15.000 € p.a.

* May be customized upon request



Training

metaphacts offers standard and custom training for knowledge graphs, metaphactory and the ResearchSpace platform.

	Standard Training	Custom Training
Preparation of training		
Deployment of training software on customer site or in metaphacts cloud	✓	✓
Custom setup with customer data/ontology, project / customer specific examples	×	✓
Slide set	Standard	Customized
Duration	1 - 2 days	Up to 5 days (incl. preparation)
Content		
Introduction to semantic technologies and knowledge graphs (RDF, OWL, SPARQL)	✓	✓
Overview metaphactory / ResearchSpace platform	✓	✓
Semantic search	✓	✓
Visualization of semantic data and knowledge graphs	✓	✓
Template development and customization	✓	✓
Configuration and administration metaphactory / ResearchSpace platform	✓	✓
User management, access control, security	×	✓
Form-based authoring	×	✓
Development of own components	×	✓
Hands-on exercises	Standard exercises	Custom exercises
Fee	2.500 €	5.000 €