

SCHEDULE 2: THE REQUIREMENTS

All functional assumes the use of CIDOC CR as the main ontology schema for data Components should be documented

Iteration 1 – Search improvements, Forms

Iteration completion date: May 31, 2017

	Title	Description	Contract Status	Notes	Est.
S1	Refinement of free text Search	The current configuration of free text needs more configuration. The user should have access to keywords and phrase construction similar to those available to systems like Google. For example, the ability to specify exact phrase tokens using quote marks, the ability to use AND, OR and NOT. The work should be informed by some example queries and an optimal solution using solr put into place.	Fixed	Requires representative queries. Metaphacts will support standard solr search syntax + improve default search configuration	5
S2	Search performance	<ul style="list-style-type: none"> • performance: the search experience should allow for fast response times • consistency: the search components should deliver consistent results (e.g. as of now, due to the LIMIT clause, search results may be inconsistent with result visualisation) • relaxation of limitations: The Facet system needs to be optimised for higher numbers of results. E.g., in calculating facets, we currently compute facet values for only 1000 results. This number should be relaxed as far as possible/reasonable <p>As a starting point, we</p>	Fixed (Time boxed)		10

		<p>perform analysis to better understand the current bottlenecks and limitations.</p> <p>Consequently, we will address these bottlenecks. Instruments for optimisation include:</p> <ul style="list-style-type: none"> • Ordering of results for consistency • More intelligent pagination (client side) to only fetch data as it is displayed (lazy loading) • Streaming pagination as results come in / process results as streamed • User feedback when query timeout is reached • Facets – lazy loading of facet counts, and caching of results. • Calculate certain things only on demand (as user requests), e.g. calculate facets only for FC explicitly selected by the user. 			
S3	Preparation for the incorporation of the FORTH alignment work	<p>The work being done by FORTH requires incorporation into the system.</p> <p>Incorporation involves:</p> <ul style="list-style-type: none"> • modification of the ETL process to materialise alignments • extension of UI components (tree selector and faceting) to support multiple, aligned thesauri <p>On the UI side, we will support a global selector that allows to specify</p>	Fixed	<p>The current BM thesauri is used. A user selects a term and the narrower terms are automatically selected.</p> <p>In the alignment system it would also be selecting narrower and exact matches that have been aligned.</p> <p>Facets show the values of the thesaurus that has been selected by the user</p>	10

		<p>whether to use local or global (aligned) thesauri.</p> <p>If thesauri have been aligned to Getty and Getty is the displayed thesauri then the results should include all terms that are narrower than the Getty term that have been assigned as narrower by the alignment tool.</p> <p>Initial test can be done by</p> <p>Loading in the Getty thesauri Configuring some alignments to simulate the results of the tool.</p> <p>Checking results.</p>			
F1	tree view picker (e.g. thesauri)	We need to extend the field configurations to support hierarchical selection, need to expose tree selector as UI component for forms.	Fixed	This should amongst other things support the use case of selecting and adding terms into a hierarchical thesaurus in forms, but be schema agnostic.	5
F2	default value in field catalogue	We need to extend the field configurations to support default, need to expose default values in all input components. (Including validation of default values against datatypes)	Fixed		2
F3	Adding Images	We will only manage references / URLs of images.	Fixed	We will only manage references / URLs of images.	2
F4	Asserted images to appear on the canonical search	The user can source images and add them quickly (with metadata and IP information, e.g. via existing file upload). This asserted data can then be used in the mini template displays.	Fixed		1

Iteration 2 – Data Processing, Public Endpoint
Iteration Completion Date: June 30, 2017

	Title	Description	Contract Status	Notes	Est
D1	FRs and Display Labels	<p>Refinement of Display labels and FRs</p> <ol style="list-style-type: none"> Multiple copies of the same result are shown in the results pane e.g. if you add both to the clipboard then only one appears. E.g Gaifu kaisei 凱風快晴, paper, Edo Period (1906,1220,0.525). This may be caused by objects having multiple titles. The system therefore needs to ensure that one display title is created – per object or that only one display title is displayed. Additional FRs for data not currently searchable and new “type” searches based on separating out other authorities. Some things are currently not searchable and require the design of new FRs. 	Fixed (Time Boxed)	Should be documented to assist additional and future work. E.g. issues, lessons learnt, approach, best practices.etc)	10
D2	Filtering on CRM types	<p>It should be possible to filter the results by specific CRM entities and provide user defaults.</p> <p>For example, a dataset contains THINGS that are both Information carriers (like photos) and Man-made object like pots. The user can specify a type, but more broadly the user should be able to declare the CRM level of Thing. E.g. Give me just Information Carriers (Photos, Pictures etc.)</p> <p>We will implement this using existing search functionality by introducing new FCs/FRs.</p>	Fixed		2

D3	ETL – creating the bridge between data inputed through the Forms and the RS Search component	<p>E.g. The Hokusai Forms instance is setup to allow entering raw data in the forms. The project will also import RDF data from other museums to be edited and enhanced using the created forms. To support the RS search the ETL needs to be run and new FRs specific to the new datasets added. As data is continuously updated, the ETL will need to be rerun.</p> <p>Ideally, this process should be automated with minimal down-time. Currently, this would be done in a series of steps: create new script to extract the raw-data from the forms instance, run the ETL and upload the data. This means the triplestore should be in "freeze" mode until the ETL finishes.</p>	Fixed (Time Boxed)	<p>UI to see and change parameters and be able to run it and see errors.</p> <p>This should have more user orientated documentation</p>	3
D4	Sample Dataset	Creation of a sample dataset that takes records from across the range.	Fixed (Time Boxed)	The aim would be to select records across a whole dataset and construct the associated authorities to make it whole for RS purposes. Should work across any CRM dataset	3
D5	Endpoint	To complete the endpoint work (templates), test and prepare publishing a new instance of the BM endpoint. To make optimisations and add additional agreed enhancement within the allocated time, e.g. specific http: API calls.	Fixed (Time Boxed)	Operation and maintenance of the endpoint is responsibility of the BM	5

Iteration 3 – Setup and Distribution

Iteration Completion Date: July 31, 2017

#	Title	User Story	Contract Status	Notes	Est.
SD1	LDAP Implementation	<p>The project requires a server/client component that is part of the standard install that provides LDAP functionality for changing passwords, adding and removing groups, adding and removing accounts from groups and user profile information.</p> <p>Outputs -</p> <ul style="list-style-type: none">Provisioning of working open source LDAP server as docker image (Apache)Provisioning of password manager (PWM) server as docker image.Documentation of best practices for setup within ResearchSpace platform	Fixed		5
SD2	Separate out databases	<p>We need to be able to separate read-only repository (with canonical data) from read/write repository with user data (e.g. forms, assertions, annotations, clipboard etc.)</p>	Fixed	<p>The idea is to have component similar to <semantic-form>, which takes list of field IRIs (can be also specified with sparql query using helper functions). Inside this component</p>	10

				<p>one will be able to reference fields to specify the template that should be used for visualization.</p> <p>We are also going to support configurable default visualizations that will be applied if one omit custom template.</p> <p>(when we discussed this previously the effort estimate was less. Need to understand the larger effort)</p>	
SD3	Federated Setup	<p>Support search to work on multiple repositories in a federated setting. In the most simple case, search will be directed towards one selected repository (targeted federation), in a more advanced case one could attempt search over multiple selected federation members.</p> <p>(All other components, including templated visualisation, are excluded from this initial work. I.e. visualisation would need to be taken</p>	Fixed	<p>It should be noted that in a federated setting, the query performance of a centralised integration cannot be guaranteed.</p>	15

		care of by the respective data publisher of the data set)			
--	--	--	--	--	--