

SCHEDULE 2: THE REQUIREMENTS

All functional assumes the use of CIDOC CRM as the main ontology schema for data

Components should be documented

Title	Description	Est	Comments
Data Cleanup	Improve data quality (time boxed)	2	Before data loading, possible data quality issues will be addressed
AWS infrastructure	Configuration of VPC, Security Groups, S3	2	Possibly using CloudFormation scripts
Neptune setup	Setup and configuration of Neptune endpoint	1	Possibly using CloudFormation scripts
Data Loading	Load collection data with both BM and extended Hokusai data into the Neptune	2	
SPARQL Endpoint	Setup of the ResearchSpace SPARQL endpoint on top of Neptune.	1	
Monitoring	Setup endpoint monitoring using AWS CloudWatch	2	
Support for production	Analysis of different pricing, fail-over,	1	

Title	Description	Est	Comments
use	backups models		
Stress testing	Stress test ResearchSpace SPARQL endpoint under load and adjust the setup accordingly	10	<ul style="list-style-type: none"> • Use realistic ResearchSpace queries generated by ResearchSpace components (in particular search component) for stress testing. • Compare query execution results between Blazegraph and Neptune to asses difference in the semantics of query evaluation. <p>Outcomes:</p> <ul style="list-style-type: none"> • Script for running stress tests based on the list of given queries. • Stress tests report. • Result similarity report. • Estimates for adjustments needed to run full ResearchSpace platform on the Neptune.

The work should be completed by 31 July 2018.