D2.3 Report on modules, transformations and API's for framework

Grant Agreement nr	780470
Project acronym	SAUCE
Project start date (duration)	January 1st 2018 (36 months)
Document due:	30th June 2018 (One month extension)
Actual delivery date	31st July 2018
Leader	Peri Friend
Reply to	William Greenly – wmg@dneg.com
Document status	Final Version

Project funded by H2020 from the European Commission

Project ref. no.	780470
Project acronym	SAUCE
Project full title	SAUCE
Document name	D2.3 Report on modules, transformations and API's for framework
Security (distribution level)	Consortium
Contractual date of delivery	30th June 2018
Actual date of delivery	31st July 2018
Deliverable name	D2.3 Report on modules, transformations and API's for framework
Type	Report
Status & version	2.4
Number of pages	24
WP / Task responsible	2 / DNEG
Other contributors	BUT
Author(s)	William Greenly - DNEG
EC Project Officer	Ms. Cristina Maier, Cristina.MAIER@ec.europa.eu
Abstract	This document outlines provides analysis of asset descriptors, proposes a new generalised asset descriptor and describes a search and transformation framework for discovering and reusing assets
Keywords	transformations, smart assets, asset descriptor, USD, metadata, ontologies, taxonomies, seach, classifiers
Sent to peer reviewer	Yes
Peer review completed	Yes
Circulated to partners	No
Read by partners	No
Mgt. Board approval	No

Document History

Version and date	Reason for Change
2.0 25-07-18	document created by William Greenly
2.1 25-07-18	Formatting updated by Peri Friend
2.2 26-07-18	Typos fixed – Review turned on by Peri Friend
2.3 30-07-2018	Revised version and final draft produced in response to Peer Review
2.4 23-10-2018	Improved document metadata as advised in quality assurance report

Table of Contents

1 5

2 Error! Bookmark not defined.

2.1 Error! Bookmark not defined.

2.2 Error! Bookmark not defined.

2.3 Error! Bookmark not defined.

2.4 Error! Bookmark not defined.

3 Error! Bookmark not defined.

3.1 Error! Bookmark not defined.

3.2 Error! Bookmark not defined.

3.3 Error! Bookmark not defined.

3.4 Error! Bookmark not defined.

4 Error! Bookmark not defined.

4.1 Error! Bookmark not defined.

4.2 Error! Bookmark not defined.

4.3 Error! Bookmark not defined.

4.4 Error! Bookmark not defined.

4.5 Error! Bookmark not defined.

4.6 Error! Bookmark not defined.

5 Error! Bookmark not defined.

5.1 Error! Bookmark not defined.

5.2 Error! Bookmark not defined.

5.2.1 Error! Bookmark not defined.

5.2.2 Error! Bookmark not defined.

5.3 Error! Bookmark not defined.

5.3.1 Error! Bookmark not defined.

5.3.2 Error! Bookmark not defined.

5.4 Error! Bookmark not defined.

5.4.1 Error! Bookmark not defined.

5.4.2 Error! Bookmark not defined.

5.4.3 Error! Bookmark not defined.

5.4.4 Error! Bookmark not defined.

5.5 Error! Bookmark not defined.

5.5.1 Error! Bookmark not defined.

5.5.2 Error! Bookmark not defined.

5.5.3 Error! Bookmark not defined.
5.6 Error! Bookmark not defined.
5.7 Error! Bookmark not defined.
5.7.1 Error! Bookmark not defined.
5.7.2 Error! Bookmark not defined.
5.7.3 Error! Bookmark not defined.
6 Error! Bookmark not defined.
7 Error! Bookmark not defined.
7.1 Error! Bookmark not defined.

1 EXECUTIVE SUMMARY

This document consists of 3 main parts. In the first part we provide background to the project, a description of the problem domain, the main objectives, and issues and use cases we are trying to address.

We then proceed to address these issues by firstly assessing the different asset descriptors that are commonly used in production pipelines, both for features and reusability. We discuss the advantages and disadvantages of each descriptor from two perspectives, firstly in the context of ingestion and search, and secondly from a reusability point of view. We proceed to identify the intersections and commonalities of each descriptor and discuss other considerations around the interchangeability vs 'last mile' aspects of each descriptor, along with the important data and information not included in an asset descriptor which is relevant for search purposes.

Having concluded that the descriptors formats, whilst suitable for reuse in a production pipelines, are not suitable for search, query, projection or aggregation, we then propose a generalised descriptor and canonical data format along with a core set of domain ontologies, which all the descriptors can be transformed to. We demonstrate that it is extendable and relatable across real world domains not covered by descriptors, provides a means for ingestion and search, and enables us to deliver on a number of other use cases such as provenance (keeping assets up to date), metadata, classification and copyright.

Finally, we provide an overview of the conceptual and technical solution and framework which enables us to deliver on the functional use cases as well as non-functional use cases and requirements, whilst at the same time being decoupled and discrete. We provide a detailed conceptual and technical overview of each of the core components of the framework, highlighting key API's, conventions and architectural patterns for each, along with the tooling and software development lifecycle required to deliver them.