Platform Interoperability Guidelines

November 20, 2017

Deliverable Code: D5.6
Version: 1.0
Dissemination level: Public

Final version of the guidelines for infrastructure interoperability structured into sets that target the stakeholder groups (providers of content and software resources)

. . .

Document Description

D5.6 – Platform Interoperability Guidelines

WP5 – Interoperability Framework	
WP participating organizations: ARC, USFD, UNIMAN, AK, UoG, GRNET	
Contractual Delivery Date: 9/2017	**Actual Delivery Date:** 11/2017
Nature: Report	**Version:** 1.0
Public Deliverable	

Preparation slip

	Name	Organization	Date
From	Penny Labropoulou	ARC	16/11/2017
	Dimitris Galanis	ARC	
	Angus Roberts	USFD	
	Matt Shardlow	UNIMAN	
	Giulia Dore	UoG	
	Thomas Margoni	UoG	
	Byron Georgantopoulos	GRNET	
	Panagiotis Zervas	AK	
	Pythagoras Karampiperis	AK	
	Richard Eckart de Castilho	UKP-TUDA	
	Antonis Lempesis	ARC	
	Lucas Anastasiou	OU	
Edited by	Penny Labropoulou	ARC	20/11/2017
	Dimitris Galanis	ARC	
Reviewed by	Mappet Walker	FRONTIERS	20/11/2017
	Marta Villegas	BSC	
Approved by	Androniki Pavlidou	ARC	20/11/2017
For delivery	Mike Hatzopoulos	ARC	20/11/2017

. . .

Document change record

Issue	Item	Reason for Change	Author	Organization
V0.1	Draft version	Initial version sent for comments	Penny Labropoulou	ARC
v0.2	Draft version	Version with comments from reviewers	Penny Labropoulou	ARC
v1.0	Final version	Final version	Penny Labropoulou	ARC

openM1N7ED

. . .

Table of Contents

. . .

Table of Figures

openM1N7ED

. . .

Disclaimer

This document contains description of the OpenMinTeD project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the OpenMinTeD consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 28 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors. (http://europa.eu.int/)	European Commission
OpenMinTeD is a project funded by the European Union (Grant Agreement No 654021).	

openM1N7ED

· · ·

Acronyms

API	Application Programming Interface
LR	Language Resource
NLP	Natural Language Processing
ML	Machine Learning
OA	Open Access
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OKFN	Open Knowledge Foundation
OMTD	OpenMinTeD
OWL	Web Ontology Language
PDF	Portable Document Format
RDF	Resource Description Framework
REST	Representational State Transfer
RI	Research Infrastructure
SKOS	Simle Knowledge Organization System
SOAP	Simple Object Access Protocol
TDM	Text and Data Mining
VM	Virtual Machine
WP	Workpackage
XML	Extensible Markup Language
XSD	XML Schema Definition

. . .

Publishable Summary

The current deliverable brings together the final version of the OpenMinTeD platform guidelines, building upon the previous version *D5.5 - Platform Interoperability guidelines (1st edition)*.

The guidelines present a user-friendly specification to empower interoperability between content and software resources, especially in the framework of the OpenMinTeD platform. It is based on input from

- *D5.4 - Interoperability Requirements Reports* (and its previous versions *D5.2* and *D5.3*) that includes the interoperability specifications set for OpenMinTeD*,
- *D6.1 - Platform Architectural Specification* that describes the architecture and functions of the OpenMinTeD platform, and
- the data model adopted by OpenMinTeD for describing resources involved in TDM and implemented in the OMTD-SHARE metadata schema.

The deliverable presents the work and methodology according to which the guidelines have been created, while the actual guidelines are published online at https://guidelines.openminted.eu.

Four guidelines have been created, targeting respectively the providers of publications, corpora, TDM software (applications and components) and ancillary knowledge resources. The specifications determine technical (e.g. data representation formats, transfer protocols), legal and documentation (metadata) issues. Two levels of compliance are foreseen, corresponding to *mandatory* and *recommended* specifications, allowing for a gradual adoption by stakeholder groups.

. . .

1. Introduction

OpenMinTeD enables the creation of an infrastructure that fosters and facilitates the use of text and data mining (TDM) technologies in the scientific publications world and beyond, by both application domain users (i.e., scientists, technicians, etc.) and text mining experts. OpenMinTeD builds upon existing tools and text mining platforms. It aims at rendering them discoverable through the OpenMinTeD registry, and interoperable through the interoperability layer, also based on existing and emerging standards and best practices.

Deliverable D5.6 presents the guidelines that interested parties must follow in order to be compatible with OpenMinTeD interoperability specifications. To serve better the needs of the target stakeholder groups and the peculiarities of each resource type, separate guidelines are available per resource type and provider group.

Deliverable *D5.5 - Platform Interoperability guidelines (1st edition)* included a short presentation of the OpenMinTeD platform and the target audience as well as the main parameters that have influenced the structure and contents of the guidelines. Readers who are not familiar with OpenMinTeD may refer to that version for an overview.

The present version reports on the work that has been done since the previous phase. The updated version of the guidelines themselves is online at: https://guidelines.openminted.eu.

. . .

2. *Structure and contents of the guidelines*

The guidelines are published online in the form of a book (with GitBook) so that they can be easily updated to reflect changes and additions in the registry and the interoperability specifications.

The introductory chapter is devoted to the general presentation of the OpenMinTeD platform and the way it has been designed and configured to integrate resources from connected content and software sources and deploy them for creating and running TDM processes (cf. *Figure 1*).
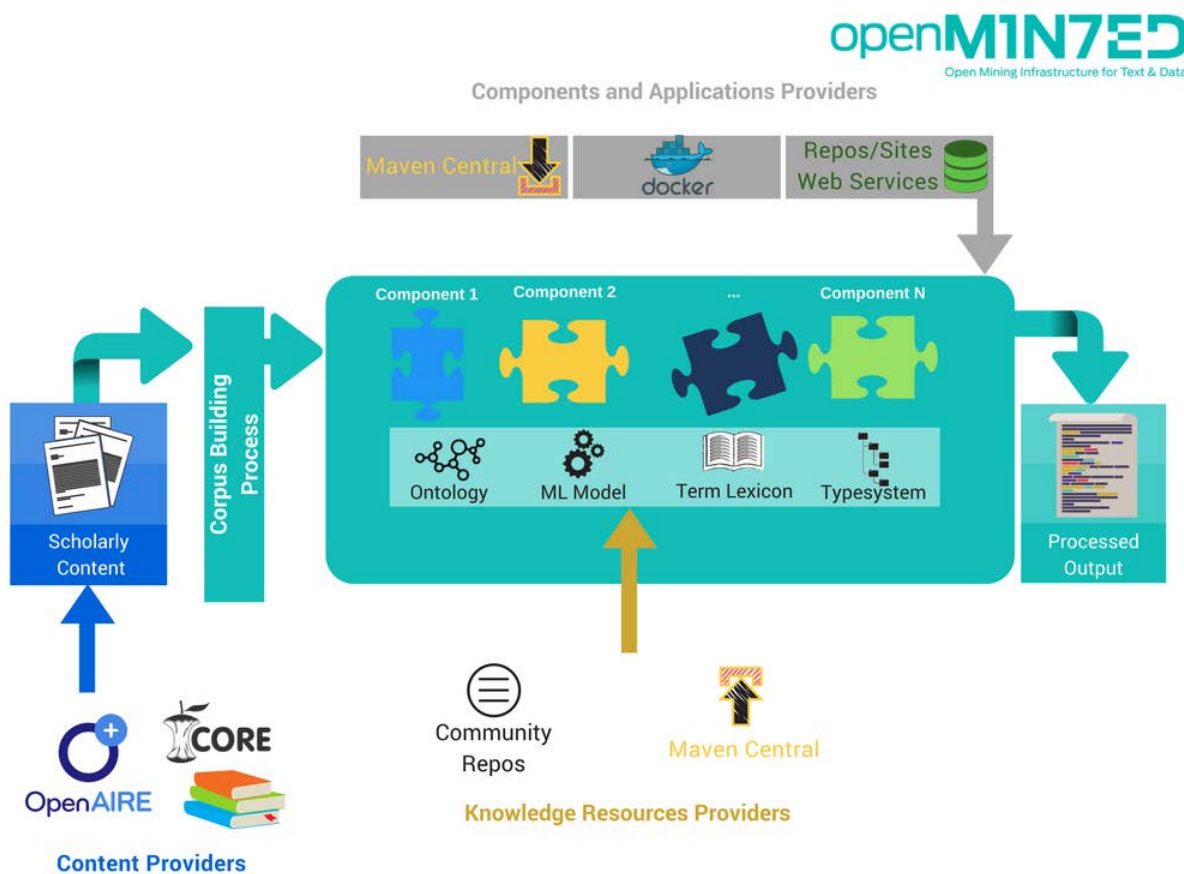
Figure 1. Overview of the OpenMinTeD platform

. . .

The general principles of the OMTD-SHARE metadata schema (cf. *Figure 2*), one of the building pillars of the platform as it provides the interoperability glue between all the resources involved, are presented in a separate chapter.
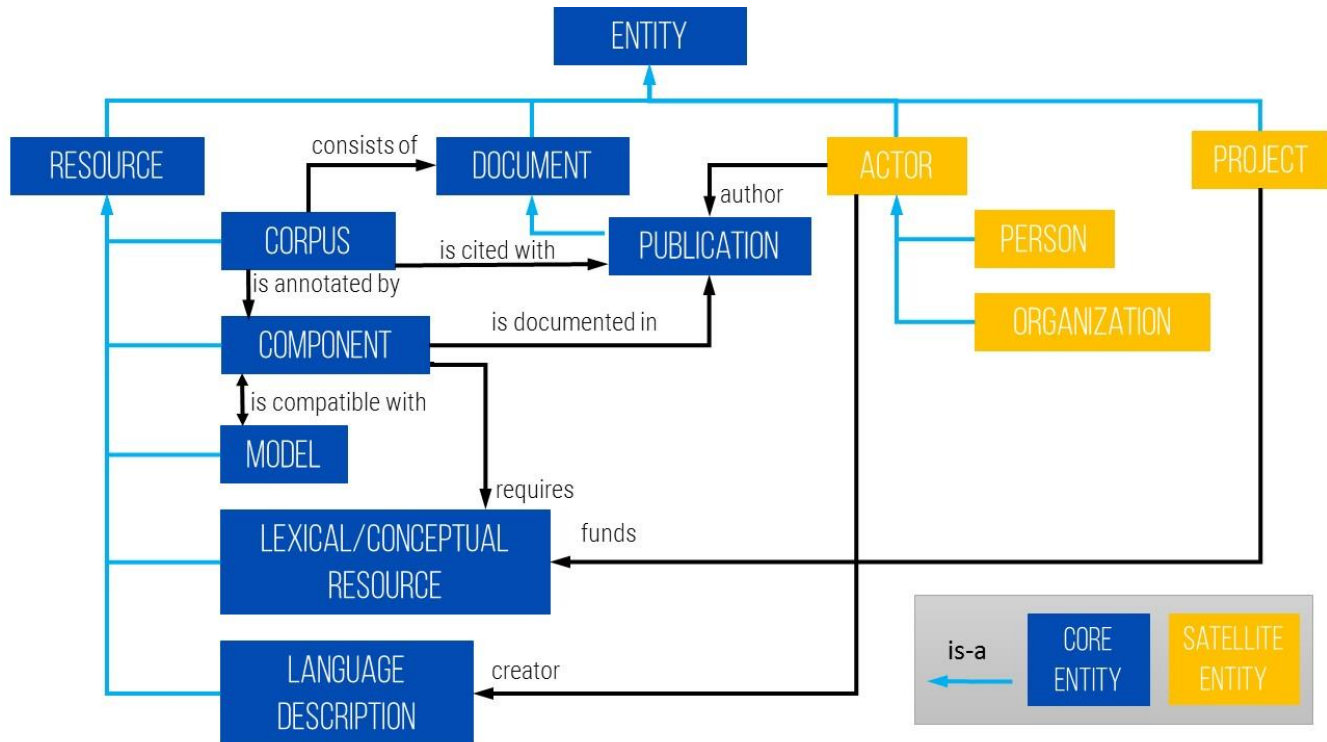

Figure 2. Overview of the OMTD-SHARE data model

The main part of the guidelines is divided into four main chapters corresponding to the three major distinctions of resources involved in TDM processes:

- *content resources to be mined*, i.e. scholarly publications,
- *TDM (-related) software*,
- *ancillary (knowledge) resources* used for the operation of the software (e.g. annotation schemas, linguistic tagsets, lexical or ontological resources used for annotating the resources to be mined, machine learning models)

and one more for

- *corpora,* i.e. collections of texts, as they can be used either as an ancillary resource (e.g. for training

. . .

models) or as a resource to be mined (e.g. subsets of the Open Access publications that are offered through OpenMinTeD selected by user queries).

Each of these chapters contains the following information:

- a brief introduction, specifying the resources expected, potential sources, minimal requirements for the contributions;
- packaging and registering instructions for the OpenMinTeD registry;
- technical and metadata requirements that empower interoperability;
- for each resource type, an overview of the OMTD-SHARE metadata schema (minimal level[1]), with definitions, explanations, recommended usage and mappings to other widespread metadata schemas;
- further instructions per type of contributors or resource type/subtype where required.

Finally, there are two GitBook functionalities that are used in the Guidelines to better address user needs:

- a Glossary of the terms that are used in the guidelines[2] with appropriate definitions is published as a separate chapter; in addition, the terms inside the texts are highlighted with a dotted underline style and when users move over them the definition is displayed;
- discussion threads can be opened by users for each paragraph of the guidelines; instructions on this feature are provided in a chapter calling for contributions. This is an important support feature in the improvement of the guidelines as it allows taking into account user recommendations.

---

[1] The OMTD-SHARE schema classifies elements into three levels of optionality: *mandatory* elements, which are necessary for discovering resources and for triggering operations between content and software components; *recommended* elements that can help the current or future use of the resource, or useful information that providers have not yet standardized; and *optional* elements with additional information related to the lifecycle of a resource. The **minimal level** includes only mandatory and recommended elements.

[2] This is part of the OMTD glossary published in the form of a SKOS vocabulary.

. . .

3. *Work progress and current status*

The main changes that have been made since the previous version include:

- updates of the contents mainly in the chapter for providers of software (applications and components) following the latest relevant technical decisions
- update of the documentation on the metadata schema (minimal version) in accordance with the latest release[3]
- addition of example metadata records per resource type
- a re-arrangement of the material, stylistic improvements and addition of images and further explanatory texts to enhance readability and make the guidelines more appealing
- addition of the chapter on user contributions.

The guidelines are used as part of the documentation for the OpenMinTeD Open Call I for Content (https://openminted.bsc.es/index.php/omtd-open-tender-calls-overview/) and also for the upcoming Call II for Services. Their use by various experts outside the OpenMinTeD consortium in response to the Open Calls will provide valuable feedback on their usability and result in updates if required.

---

[3] The full OMTD-SHARE schema is documented at: https://openminted.github.io/releases/omtd-share/