



# On-line Support Knowledge Base

April 13, 2018

**Deliverable Code:** D3.6

**Version:**1.0 – Final

**Dissemination level:** Public

The goal of the On-line Support Knowledge Base report is to provide the consortium with the technical, legal and policies expertise directories that were created in WP3 with the aim offer to stakeholders expertise guides on important Text and Data Mining components.



H2020-EINFRA-2014-2015 / H2020-EINFRA-2014-2  
Topic: EINFRA-1-2014  
Managing, preserving and computing with big research data  
Research & Innovation action  
Grant Agreement 654021



# Document Description

## D3.6 – On-line Support Knowledge Base

WP3–Support and Training	
<b>WP participating organizations:</b> ARC, University of Manchester, UKP-TUDA, INRA, EMBL, Agro-Know I.K.E., OU, EPFL, CNIO, USFD, GESIS, GRNET, Frontiers, UoS, LIBER	
<b>Contractual Delivery Date:</b> 03/2018	<b>Actual Delivery Date:</b>
<b>Nature:</b> Report	<b>Version:</b> 1.0 Final
<b>Public Deliverable</b>	

### Preparation slip

	Name	Organization	Date
<b>From</b>	Petr Knoth	OU	26/03/2018
	Nancy Pontika	OU	14/03/2018
	Lucas Anastasiou	OU	22/03/2018
	Martine Oudenhoven	LIBER	22/03/2018
<b>Edited by</b>	Nancy Pontika	OU	30/03/2018
	Martine Oudenhoven	LIBER	09/04/2018
<b>Reviewed by</b>	Andrea Zielinski	GESIS	28/03/2018
	Panagiotis Zervas	AgroKnow	30/03/2018
<b>Approved by</b>	Androniki Pavlidou	ARC	30/04/2018
<b>For delivery</b>	Mike Hatzopoulos	ARC	30/04/2018

### Document change record

Issue	Item	Reason for Change	Author	Organization
V0.1	Draft version	Initial version sent for comments	Nancy Pontika	OU
V1.0	First version	Incorporating reviewers' comments	Nancy Pontika	OU



## Table of Contents

<b>1. INTRODUCTION.....</b>	<b>5</b>
<b>2. EXPERTISE DIRECTORIES .....</b>	<b>6</b>
<b>2.1 TECHNICAL EXPERTISE DIRECTORY .....</b>	<b>6</b>
<b>2.2 LEGAL AND POLICY EXPERTISE DIRECTORY .....</b>	<b>7</b>



# 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/>)



OpenMinTeD is a project funded by the European Union (Grant Agreement No 654021).



# Acronyms

TDM	Text and Data Mining
WP	Work Package
OMTD-KB	OpenMinTeD Knowledge Base
DoW	Description of Work



## 1. Introduction

The purpose of WP3 is to assist both the project partners but also the project stakeholders with the training related to Text and Data Mining (TDM). This WP consists of three tasks and this report relates only to Task 3.2 “Support services.” The goal of the on-line support expertise directory report is to provide the project stakeholders with expertise directories in two important components of Text and Data Mining (TDM) and create two expertise directories in the following topics:

1. Technical Expertise Directory
2. Legal & Policy Expertise Directory

The work on the Technical expertise directory was led by The Open University, while LIBER lead the Legal and Policy Expertise Directory.

Each one of these directories provides structured information in the aforementioned topics and addresses important components of TDM. A brief summary is presented above, while more extended descriptions are included in the following subsections:

### **Technical Expertise Directory:**

The purpose of this directory is to demonstrate how scientific content from publishers can be accessed. It illustrates the importance for a connector between publishers and a text miners, it applies a review and evaluation of the publishers’ APIs, provides an aggregation approach to Open Access Hybrid Gold scientific publications, features a table with the publisher’s API capabilities and limitations and provides recommendations towards publishers to improve their machine accessible interfaces.

### **Policy and Legal Expertise Directory:**

The goal of the legal expertise directory is to support text and data miners by providing information on the legal aspects related to TDM. The knowledge base consists of two parts:

1. General information on the legal aspects related to text and data mining, uploaded to the OMTD-KB.
2. Detailed information that answers legal questions that may be related to interactions with the platform (related also to task 3.3).

This deliverable is classified in the DoW as “Other” and a more detailed written report about these activities will follow in D3.7.



## 2. Expertise Directories

### 2.1 Technical Expertise Directory

The focus of the technical expertise directory was the machine access to content from publisher systems. The directory was developed in Github and it has a public url <https://github.com/openminded/omtd-publisher-connector-harvester>

The Open University explored the publishers' machine interfaces and protocols (APIs) and based on this exploration it offers a detailed guide on how the open access content can be programmatically accessed for each publisher. The reasoning behind this investigation was to take full advantage of the TDM benefits and unlock hidden knowledge. In order to release the hidden knowledge, The Open University created a connector, which offers a seamless layer for accessing content across publishers, provides a generic centralized point of access to all the available resources and provides a high performance and up-to-date access interface.

The investigation was conducted in four key publishers, Elsevier, Springer, Frontiers and PLoS. Each publisher has a dedicated page in the project's Github area and the following sections are provided for each publisher:

- Publisher API
- Information provided by the publisher with regards to TDM abilities and capabilities
- Publisher's Aggregation Approach
- A features table, with information about the retrieved list of items, list of DOIs and retrieval throttling, with limitations and proposals for solutions.

Via this work which commenced in May 2017 we managed to originally release approximately 1.6 million of Hybrid Gold Open Access research papers. Since then, the connector has been continuously extracting new papers made available by one of these publishers and the total number of extracted articles reached 1.9 million (Springer 512.358, Elsevier 1.191.785, Frontiers 63.011, PLOS 176.162) in February 2018.

This work was presented in national and international conferences:

- OAI10 – CERN – UNIGE Workshop on Innovations in Scholarly Communication, 21-23 June 2017, Geneva [[poster](#)].
- Open Repositories, Brisbane, Australia, 26 -30 June, 2017 [[slides](#)].
- Repositories Fringe, Edinburgh, Scotland, 4-5 August 2017 [[poster](#)].
- FORCE 2017, Berlin, 25-27 October 2017 [[slides](#)].

The poster we created for the Publisher Connector received the best poster award in two conferences, OAI10 and the Repositories Fringe 2018. This work has also been communicated to the wider stakeholders' community via a blog post at the [London School of Economics Impact Blog](#). Currently the Open University team is composing a research article describing the research and technical process that was followed to complete this work.



## 2.2 Legal and Policy Expertise Directory

The legal and policy expertise directory provides general legal information related to text and data mining, as well as specific information that is helpful while interacting with the platform. LIBER, the Open University, the University of Glasgow and the University of Amsterdam contributed to compiling the information.

### General information

The legal and policy expertise directory provides general legal information related to text and data mining, as well as specific information that is helpful while interacting with the platform. LIBER, the Open University, the University of Glasgow and the University of Amsterdam contributed to compiling the information.

While interacting with the OpenMinTeD platform, more specific questions are expected to come up. These are answered by the legal and policy related FAQs that are available on the OpenMinTeD platform itself: <https://services.openminted.eu/support>. The FAQs include 35 questions and answers about the policies related to using the platform, including 6 general policy questions, and 13 legal questions and answers. This support page also refers to the Knowledge Base for more elaborate information related to TDM and is related to task 3.3.

Task 3.3 also covers a ticketing system in which 13 questions have been asked and answered, and a forum. The forum was initially started up as a Google group, and has had 28 threads on very specific (technical) topics. Because of technical problems of the Google group, the forum was moved to GitHub (<https://github.com/openminted/Open-Call-Discussions>). By 6 April, 14 issues had been reported on the new forum. A more elaborate description of task 3.3 is given in deliverable 3.7.

In the near future, the Knowledge Base is expected to also contain tutorials for for the use cases and the applications that are being developed by the open call winners.