

# Software Release Report

December 19, 2016

**Deliverable Code:** D7.3

**Version:** 1.1 - Final

**Dissemination level:** Public

This deliverable presents the plan of design and development of the OpenMinTeD infrastructure services for the next 12 months. At month 12 and 24 an update of this deliverable will be produced. Its intended use is mainly for technical partners to locate their software release duties in the wider context of the project software release, but also for the generic reader to get an overall picture and insight view of the technical activities.



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

**openMIN7ED**  
Open Mining Infrastructure for Text & Data



# Document Description

## D7.3 - Software Release Report

|   |                                      |
|---|--------------------------------------|
| WP7 - Platform Integration, Testing, and Deployment                                 |                                      |
| <b>WP participating organizations:</b> USFD, ARC, UNIMAN, UKP-TUDA, INRA, AK, GRNET |                                      |
| <b>Contractual Delivery Date:</b> 07/2016   | <b>Actual Delivery Date:</b> 08/2016 |
| <b>Nature:</b> Report   | <b>Version:</b> 1.1 (Final)          |
| <b>Public Deliverable</b>   |                                      |

### Preparation slip

|                     | Name              | Organization          | Date       |
|---------------------|-------------------|-----------------------|------------|
| <b>From</b>         | Mark A. Greenwood | USFD                  | 29/07/2016 |
| <b>Edited by</b>    | Mark A. Greenwood | USFD                  | 10/08/2016 |
| <b>Reviewed by</b>  | Thomas Margoni    | University of Glasgow | 08/08/2016 |
|                     | Matt Shardlow     | UNIMAN                | 09/08/2016 |
| <b>Approved by</b>  | Matt Shardlow     | UNIMAN                | 09/08/2016 |
| <b>For delivery</b> | Mike Chatzopoulos | ARC                   | 10/08/2016 |

### Document change record

| Issue | Item          | Reason for Change  | Author            | Organization |
|-------|---------------|--|-------------------|--------------|
| V0.1  | Draft version | Initial version  | Mark A. Greenwood | USFD         |
| V0.2  | For Review    | Finalised partner contributions ready for internal review                | Mark A. Greenwood | USFD         |
| V1.0  | Final Version | Incorporated reviewer feedback   | Mark A. Greenwood | USFD         |
| V1.1  | Style Change  | Updated the style to match other deliverables, content identical to V1.0 | Mark A. Greenwood | USFD         |



# Table of Contents

- OPENMINTED SOFTWARE RELEASES.....7**
  
- ONGOING PUBLIC DEVELOPMENT.....9**
- INTEROPERABILITY SCENARIO DESCRIPTIONS AND REQUIREMENT SPECIFICATIONS.....9**
- INTEROPERABILITY MAPPING CONVERSIONS.....9**
- THE OPENMINTED REGISTRY.....9**
- INFORMATION EXTRACTION FOR VITICULTURE RESEARCHERS.....10**
  
- RECENT RELATED RELEASES.....11**
- GATE 8.2.....11**
- DKPRO CORE 1.8.0.....11**
- ARGO.....12**



# 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/>)



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



# Acronyms

|     |                        |
|-----|------------------------|
| IE  | Information Extraction |
| TDM | Text and Data Mining   |



# Publishable Summary

Throughout the life of the OpenMinTeD project we are planning to release numerous versions of a number of core components that will enable interoperability between existing Text and Data Mining (TDM) frameworks. This report will be updated with these releases to provide a single reference document to point interested users to the relevant components. Our current release plan will see the first official OpenMinTeD software become available at the end of September 2016 and so this version of the report focuses on related software development.

Many, if not all, of the components we are planning to release are being developed in public and this report also gives a brief overview of this work and pointers to where to look for more information.

Alongside software developed specifically for OpenMinTeD, a number of partners are involved in the development of related TDM software. Some of this software has seen recent development inspired by work in OpenMinTeD and this report outlines that work as well for the interested reader.



## 1. OpenMinTeD Software Releases

The first formal software release from OpenMinTeD is scheduled for the end of September 2016. As such this section of the report is currently a place-holder for the documentation of future releases.

This is the current snapshot of the timeline for releases of the OpenMinTeD platform components and applications. The master timeline is hosted on the Redmine project management system, and is updated as the project's needs and architecture plans are clarified, and as they further evolve. Any proposed changes to the timeline will be discussed in the technical work packages (i.e. WP5-WP8).

Each release will follow a two-month cycle of pilot → testing → integration → deployment, so for a release target date at the end of one month the release cycle starts at the beginning of the preceding month.

| Start of Release Cycle | Release Date | Month | Milestone | Type         | Title   |
|------------------------|--------------|-------|-----------|--------------|---|
| 30/07/2016             | 30/09/2016   | 16    | MS30      | Component    | Registry                                      |
| 31/08/2016             | 31/10/2016   | 17    | MS31      | Platform     | Platform V1                                   |
| 30/11/2016             | 31/01/2017   | 20    | MS36      | Component    | User and accounting services V1               |
| 30/11/2016             | 31/01/2017   | 20    | MS37      | Applications | 1 <sup>st</sup> Community Driven Applications |
| 31/03/2017             | 31/05/2017   | 24    | MS42      | Component    | Data interoperability APIs and toolkit V1     |
| 31/05/2017             | 31/07/2017   | 26    | MS45      | Component    | User and accounting services V2               |
| 31/05/2017             | 31/07/2017   | 26    | MS46      | Component    | Workflow service                              |
| 31/05/2017             | 31/07/2017   | 26    | MS47      | Component    | Annotation presentation service               |

• • •

|            |            |    |      |              |   |
|------------|------------|----|------|--------------|---|
| 31/05/2017 | 31/07/2017 | 26 | MS48 | Platform     | Platform V2   |
| 31/07/2017 | 31/08/2017 | 27 | MS49 | Applications | 2 <sup>nd</sup> Community Driven Applications         |
| 30/11/2017 | 31/01/2018 | 32 | MS54 | Component    | Annotation editing service and crowd-sourcing plugins |
| 30/11/2017 | 31/01/2018 | 32 | MS55 | Component    | Data interoperability APIs and toolkit V2             |
| 31/12/2017 | 28/02/2018 | 33 | MS56 | Platform     | Platform V3   |
| 31/01/2018 | 28/02/2018 | 33 | MS57 | Applications | 3 <sup>rd</sup> Community Driven Applications         |





## 2. Ongoing Public Development

As discussed in the previous section, the OpenMinTeD project has not yet reached a point where it is ready to formally release any of the software being developed. It is worth noting, however, that much of the software is being developed in public and this brief section details the main development work currently taking place.

### 2.1 Interoperability Scenario Descriptions and Requirement Specifications

In this repository (<https://github.com/openminted/interoperability-spec>), we host the interoperability scenario descriptions and the interoperability requirements specification. While the scenarios are now largely stable, the interoperability requirements are still being continuously updated. This not only includes textual descriptions but also scripts that process the requirements information to generate different kinds of overviews and reports. A version is planned to be frozen and released as part of deliverable D5.2.

### 2.2 Interoperability Mapping Conversions

In this repository (<https://github.com/openminted/interoperability-mapping-conversion>), we host software to automatically aggregate and compare type systems and component descriptions. The software is regularly updated as new features are added, e.g. the crawling of metadata directly from repositories that host components. This is considered experimental software/proof-of-concept software and there is presently no release schedule. However, it is expected that parts of the code will be picked up/adapted by the task force working on the component registry.

### 2.3 The OpenMinTeD Registry

As detailed in the previous section, the first software release from the OpenMinTeD project will be the registry. Development of this component has already started, although it has not yet been moved into the OpenMinTeD GitHub repository. As per the original development and testing plan, the code is being built automatically via the project's Jenkins installation (<https://builds.openminted.eu>) and snapshots are being automatically deployed to our Nexus repository (<https://repo.openminted.eu>).



## 2.4 Information Extraction for Viticulture Researchers

Agroknow recently coordinated a hackathon which produced an API endpoint for extracting text from PDF files which could then be annotated by other endpoints. The goal is to provide end-users with an intuitive way to annotate their bibliographical resources. To this end, the FREME API (an endpoint used for entity recognition and interlinking), the AgroPortal annotator (used for the recognition and interlinking of entities against selected ontologies) and a custom SPARQL endpoint (with information about grape varieties) were used. It is envisaged that this work may be included into one of the OpenMinTeD usecase demonstrators.

See <https://github.com/agroportal/1stAgroHackathon/tree/master/hack-14> for more details.



### 3. Recent Related Releases

Outside the OpenMinTeD project itself there have recently been a number of software releases that are worth mentioning as they impact upon the project in some sense. The two releases mentioned in this version of the report are of TDM frameworks developed by project partners which will be closely integrated with the final OpenMinTeD platform.

#### 3.1 GATE 8.2

While not directly related to OpenMinTeD, GATE 8.2 has recently been released. This release sees numerous new features, and new plugins, many of which will in time be made available through the OpenMinTeD registry. More importantly, however, this release has allowed development focus to move to the next major release of GATE. There is currently no release date scheduled but most of the planned features are, in some way or another, related to OpenMinTeD. This includes changes to component metadata to allow use with the OpenMinTeD registry and a move to distributing plugins via Maven to aid automatic construction of workflows.

Development is currently taking place in the public GATE source code repository which can be found at <http://svn.code.sf.net/p/gate/code/gate/branches/sawdust2/>

#### 3.2 DKPro Core 1.8.0

The recent release of DKPro Core 1.8.0 includes several improvements that were prompted by discussion in OpenMinTeD WP 5.2 and implemented by UKP-TUDA. The new release includes a comprehensive reference documentation that is generated automatically from all kinds of sources of metadata within DKPro Core, including typesystem descriptions, component descriptions, etc. As part of this reference documentation, linkings to other type systems (GATE ANNIE, LAPPS Grid, ISOCat) were introduced as suggested in WP 5.2 WG 2. Also following a suggestion by WP 5.2 WG 2 types received a URI by which they can be linked to from type mappings. The aggregated information that is the basis for the reference documentation is also the basis for the generation of metadata according to the upcoming OpenMinTeD specification from WP 5.1 WG 1 and has certainly contributed to the process of developing this specification.

DKPro Core can be found at <https://dkpro.github.io/dkpro-core/>



### 3.3 Argo

The Argo workflow management and execution system already has a number of desirable features, which make it a powerful tool for fostering interoperability amongst developers and user communities. Built on top of the OASIS Unstructured Information Management Architecture (UIMA), Argo can easily integrate other TDM tools which use the same technology. Argo workflows can be used to perform a variety of TDM tasks and workflows can be shared across users to facilitate re-use. Argo contains an internal component registry, many of which will be exposed to the OpenMinTeD registry. Argo will also be able to draw components from the project's registry when it is available. Argo also has its own internal annotation editor and the OpenMinTeD annotation editor will be integrated with Argo's systems. Further to this, New developments in Argo contain several features which are relevant to the OpenMinTeD project. These include: improvements to the stability of the execution engine, component license management and aggregation, distributed execution supporting a number of cloud services and improved integration with web services.

Argo (available from <http://argo.nactem.ac.uk/>) is currently in beta with new features appearing frequently.