Collaboration and Liaison Plan

December 28, 2015

Deliverable Code: D2.2

Version: 1.2 – Intermediary

Dissemination level: PUBLIC

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

D2.2 – Collaboration and Liaison Plan

WP2 - Community Engagement and Sustainability	
WP participating organizations: LIBER, ARC, University of Manchester, UKP-TUDA, INRA, EMBL, Agro-Know I.K.E., University of Amsterdam, OU, EPFL, CNIO, USFD, GESIS, GRNET, Frontiers, UoS.	
Contractual Delivery Date: 12/2015	**Actual Delivery Date:** 01/2016
Nature: Plan	**Version:** 1.2 (Draft)
Public Deliverable	

Preparation slip

	Name	Organization	Date
From	Natalia Manola, Theodoros Manouilidis	ARC	20/12/2015
Edited by	Natalia Manola, Stelios Piperidis	ARC	20/12/2015
Reviewed by	Angus Roberts Byron Georgantopoulos Richard Eckart de Castilho Vassilis Protonotarios	USFD GRNET TUDA Agro-Know	7/1/2016 8/1/2016 11/1/2016 15/1/2016
Approved by	Stelios Piperidis	ARC	16/1/2016
For delivery	Mike Hatzopoulos	ARC	

Document change record

Issue	Item	Reason for Change	Author	Organization
V0.1	Initial version	Document outline	Theodoros Manouilidis	ARC
V1.0	First draft	Detailed strategy and liaisons	Natalia Manola	ARC
V1.1	Second draft	Integrated review comments/feedback	Natalia Manola	ARC
V1.2	First Delivery	Final edit	Stelios Piperidis	ARC

1. Table of Contents

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).

Publishable Summary

OpenMinTeD's objective is to establish an open and sustainable Text and Data Mining (TDM) platform and infrastructure where researchers can collaboratively create, discover, share and re-use knowledge from a wide range of text-based scientific and humanities related sources in a seamless way to advance research, promote interdisciplinary open science, and ultimately support evidence-based decision making.

This document outlines OpenMinTeD's collaboration and liaison plans. It identifies the stakeholders and ways to liaise that will enable the widest possible adoption of the infrastructure and will empower its uptake and sustainability. It establishes a clear roadmap of who it should work and liaise with, what for and what are the expected outcomes in the short and medium term.

The recurring theme for establishing synergies with similar or complementary initiatives in Europe and beyond relates to the OpenMinTeD outcomes, namely the interoperability guidelines and the platform services (content and service registry, annotation and workflow services). We focus on how to assist in the optimization of the use of resources by exchange of knowledge and technology and on how to increase the impact and awareness of TDM in an open scholarship/open science environment.

2. *Introduction*

2.1 Project Background

OpenMinTeD aspires to enable the creation of an infrastructure that fosters and facilitates the use of text and data mining technologies in the scientific publications world and beyond. It will do so by engaging with a variety of stakeholders, bringing together content providers and scientific communities, text mining and infrastructure builders, legal experts, data and computing centers, industrial players and SMEs, individual researchers and citizen scientists.

OpenMinTeD builds upon existing text mining tools and platforms, rendering them discoverable, through appropriate registries, and interoperable through an interoperability layer.

Beyond the development of the technical e-Infrastructure, OpenMinTeD aims to bring awareness of the benefits and training of text and data mining (TDM) users and developers alike and demonstrates the merits of the approach through a number of use cases identified by scholars and experts from different areas, ranging from life sciences (bioinformatics, biochemistry, etc.) to food and agriculture and social sciences and humanities related literature.

2.2 Mission and Vision

This Liaison and Collaboration plan presents an overview of with whom, and how OpenMinTeD should conduct its outreach activities over the short and medium term and establishes a clear roadmap of who it should work and liaise with.

OpenMinTeD's objective is to establish an open and sustainable TDM platform and infrastructure where researchers can collaboratively create, discover, share and re-use knowledge from a wide range of text-based scientific related sources in a seamless way to advance research, promote interdisciplinary open science, and ultimately support evidence-based decision making. Its vision and mission statements are formulated as:

Vision statement

Knowledge discovery and exploitation for all

Mission statement

OpenMinTeD initiates an infrastructural approach to open up research outputs for text and data mining, to foster knowledge discovery, and advance research and innovation within the Open Science ecosystem.

OpenMinTeD provides an interoperability layer and services to enable

1. uniform access to openly available research literature and related content, and
2. discovery, deployment and use of interoperable text and data mining resources, tools, services and workflows.

The liaison and collaboration activities focus on maximizing the impact of TDM and promote uptake of the OpenMinTeD e-Infrastucture, and optimize the use of resources by exchange of knowledge and technology. More specifically they need to address the following:

- Provide easy and homogeneous access to research publications and research related content.
  - Make the rules straightforward for content providers and content consumers alike. Promote open protocols and formats and drive their adoption throughout Europe.
  - Engage content providers to participate in the wider Open Science ecosystem. Show benefits, lower legal and technical barriers.
  - Involve legal experts to assist in IPR, licensing and contracts topics.
- Provide access to text and data mining tools and services and make them visible to a wider audience. and increase the capabilities of knowledge for all.
  - Engage TDM researchers and application developers by making their services publicly available, easily discoverable and interoperable.
  - Follow latest technology trends for cloud storage and processing. Use existing European and national e-Infrastructures.
  - Involve legal experts to assist in trusted, long-term service provision.
- Build an online community around TDM technological, organizational and legal issues
  - Engage researchers to use TDM and the OpenMinTeD e-Infrastructure and platform. Communicate TDM benefits and ease of use to a wide range of research communities, ranging from established research communities (e.g., European Strategy Forum on Research Infrastructures ESFRIs) to individual researchers that represent the long tail of science.
  - Engage policy makers and funders to promote the Open Science vision through TDM.
  - Involve legal experts to provide consultation on legal aspects of TDM and the use of the OpenMinTeD e-Infrastructure.
  - Engage with similar initiatives/e-Infrastructures from other regions of the world.

3. *Target Stakeholders*

Figure 1 illustrates the complexity of the scientific TDM domain, and shows how OpenMinTeD tries to bridge the various services and stakeholders. Among others, it brings together content providers and research communities, text mining and infrastructure builders, legal experts, data and computing centers, industrial players and SMEs, policy makers and citizen scientists.

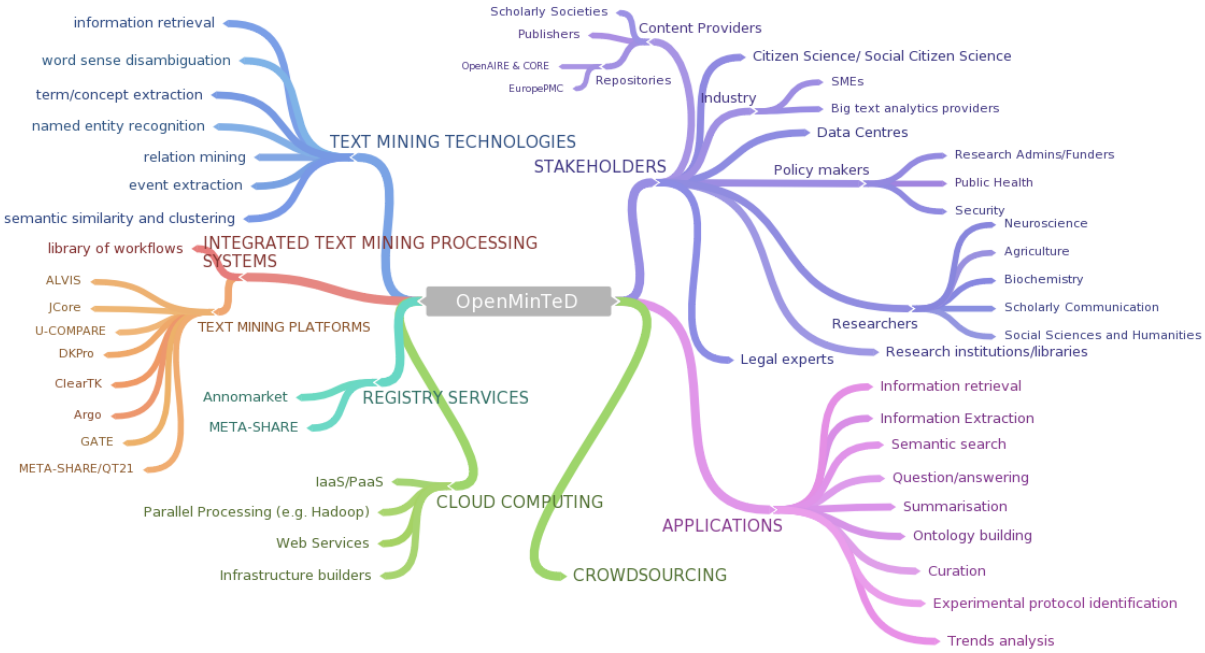

Figure 1. OpenMinTeD outreach.

OpenMinTeD primarily targets two levels of users, who often have an interchangeable role of consumer/producer (prosumer[1]):

- **End users:** researchers, curators, citizen scientists and similar type users that will consume text mining services discovered through OpenMinTeD and accessed through standard APIs. They can be novice users who seek for and use available services in order to advance their science, or they can be more advanced users (e.g., SMEs) who include text mining services into more complex research workflows. In either case, they want to achieve their tasks and get to the end result in a straightforward, seamless manner with common understanding of all aspects of use: how to provide the content and access the services, what to expect in terms of performance or quality, what are the possible legal constraints and implications;
- **Service and content providers:** Service and content providers aiming to participate in the infrastructure will provide their services or content for consumption and reuse. Service providers are text mining experts who both provide their own existing and emerging services via the platform, and also make use of the platform to extend their own work through the discovery and use of other text mining (sub-) components. Content/data providers are keen to integrate their assets in the chain of text mining ensuring that those assets will be handled as expected.

[1] https://en.wikipedia.org/wiki/Prosumer

The Collaboration and Liaison strategy plan targets a number of these stakeholders and describes specific outcomes for each one of them.

3.1.1 Repositories, publishers, scholarly societies

This group is crucial to OpenMinTeD as content provision is key in TDM. The approach methodology mainly targets to:

- encourage them to open up the content for TDM. Promote clear open licenses that allow sharing and re-using via TDM;
- illustrate the benefits (more visibility, better use by researchers and SMEs, etc.) through value added services;
- promote the use of common protocols and formats for exposing the content.

Of importance are the synergies developed with OpenAIRE, as this seems to be the prevailing e-Infrastructure for Open Access in Europe.

Who to collaborate with?	Why?	Expected Outcome
OpenAIRE	OpenAIRE is an aggregator with outreach to many different OA repositories and journals. It has developed, uses and exposes a variety of text mining services and many of its services rely on them. OpenAIRE's extensive network is an excellent gateway to reach out to OA repository managers and researchers alike and make them TDM aware. OpenAIRE's Zenodo is an additional content provider.	Access of primary content. OpenAIRE has the infrastructure to store publications and has already started making agreements with providers. Can play the role of content broker, both on legal and technical aspects. Additionally it can be a case study for multi-lingual TDM.
Jisc/CORE	CORE is an aggregator of content from UK repositories that also stores full texts.	Broker to provide content to OpenMinTeD users (provided that licenses or consents are in place).
HAL/ISTEX	HAL is the French national repository platform that contains full texts. ISTEX is an ongoing national project that aims to provide to higher education and research community online access to retrospective collections of scientific literature in all disciplines by setting up a national document acquisition policy covering journal archives, databases, texts corpora etc.	Primary content to be accessible through common APIs and access rules.

Who to collaborate with?	Why?	Expected Outcome
BASE	A global aggregator (metadata only) with outreach to more than 84 mi publications.	Promotion of common protocols and formats to OA repositories/sources.
FAO/Agris	Global publication aggregator for the agriculture scientific domain. Contains mainly metadata and often full text.	Mediator to agriculture thematic literature repositories. Promotion of common protocols, formats, APIs.
Paperity	A multidisciplinary aggregator of Open Access journals and papers (900K papers)	Promotion of common protocols and formats to OA repositories/sources.
arXiv	Domain discipline repository, its content downloadable from Amazon's S3 cloud services. The arXiv team has shown interest for TDM services for own use. OpenAIRE mirrors its content, so no immediate need to act as a content provider.	Use of TDM services to enrich metadata.
bioRxiv	Domain discipline repository for life sciences.	Access to primary content, via the promotion/use of common protocols and formats.
PMC Europe	Domain discipline repository, its OA content downloadable via ftp. PMC Europe provides on- and offline TDM services to its users. OpenAIRE mirrors its content, so no immediate need to act as a content provider.	Access to primary content, via the promotion/use of common protocols and formats.
DOAJ	Aggregator/catalogue of OA journals.	Promotion/use to its constituency of common protocols and formats.
DataCite	Aggregator of data repositories which many times include text content.	Promotion/use to its constituency of of common protocols and formats
Frontiers PLOS	OA publisher keen to open up content for TDM.	Access to primary content, via the promotion/use of common protocols and formats.

Who to collaborate with?	Why?	Expected Outcome
Wiley	Publisher that may be interested to open up for TDM and does not yet have in house TDM technology	Promotion/use of common protocols and formats.
Mendeley ResearchGate	Researcher network with full text content.	Promotion/use of common protocols and formats.
CrossRef	CrossRef already has issued some protocols and APIs for TDM. It is important to align efforts and bring all content sources (repositories and publishers) to use the same technological solutions.	Promotion/use of common protocols and formats.
ACM	ACM is looking into ways to enhance their content and provide value added services to their users.	Use of common services. Expose content via common protocols and formats.
Figshare	All data-type repository with full text content. Figshare is currently expanding its business model to serve universities, potentially acting as a broker.	Use of common services. Expose content via common protocols and formats
Europeana	Aggregator of cultural heritage related content.	Promotion of common standards. Broker for scientific related/used content (e.g., newspapers).
EPO	The European Patent Office provides search services and APIs to allow TDM on patents.	Access to primary content. Promotion of APIs and guidelines. Transformation services.
USDA NAL	The National Agricultural Library of the US Dept. of Agriculture is one of the major agricultural research publishers globally. They invest in new technologies related to research publications (PubAg) and data (Ag Data Commons)	Explore opportunities for collaboration; e.g. identify interest to test and adopt TDM services on top of their content (publications and data).

3.1.2 Text mining and language researchers

This group will have a two-fold role:

- register their resources (services, tools and language resources) to the OpenMinTeD infrastructure to expose them to a broad range of users;
- use the OpenMinTeD guidelines and platform services to reach out to content and showcase their services to reach out to specific domain discipline research communities that are interested in mining scientific literature.

The OpenMinTeD consortium already includes a number of top-notch NLP labs, each bringing in their tools and services, which will be aligned to the new guidelines and adapted to be part of the platform or applications developed on top. In addition, the consortium will look out to additional text mining research teams through the following venues:

Who to collaborate with?	Why?	Expected Outcome
META-SHARE	META-SHARE is a network of language resources which already contains a registry of resources, similar to the one proposed in OpenMinTeD (meta-share.eu/org, qt21.metashare.ilsp.gr) META-SHARE has also looked into and come up with licensing schemes that may be useful to OpenMinTeD.	Technology and resources share/re-use. Outreach to NLP labs around Europe and the world. Promotion/use of common guidelines.
CLARIN	CLARIN EU is a pan-European infrastructure setting the foundations for language resources and tools documentation, persistent identification, preservation and lawful sharing. In addition, CLARIN EU aims at fully deploying a single sign on (SSO) policy based on SAML2.0.	Outreach to the CLARIN research community for common protocols and formats. CLARIN can also be a mediator for language resources.
GateCloud	GateCloud is TDM infrastructure in the cloud, based on the earlier FP7 AnnoMarket project. It operates a registry of TDM services, run on Amazon's cloud. It has an SME and researcher user base which can be used in OpenMinTeD.	Technology and resources sharing. Promotion and use of common guidelines. Outreach to SMEs.
Know-Center	Among other services, Austria's Know-Center has developed Sensium, a scalable data mining and analysis platform. Know-Center has a team of young scientists active in Open Science.	Technology share. More outreach, more services.

Who to collaborate with?	Why?	Expected Outcome
BioCreative	The BioCreative evaluation challenge consists of a community-wide effort for evaluating text mining and information extraction systems applied to the biological domain.	Use of platform services & APIs to retrieve content. Outreach to contesters.
FREME	A H2020 project that provides an open framework of e-services for multilingual and semantic enrichment of digital content.	Sharing of tools/ services via common APIs.
NCBO Web Services	The BioPortal provides services to allow access to biomedical terminologies and ontologies, lexicons, controlled terminologies and ontologies that can describe and index the contents of online data sets (data annotation).	Resource metadata and description. APIs to retrieve biomedical ontologies and lexicons.

3.1.3 Cloud and Data infrastructure

A key part of the TDM e-infrastructure is the use of cloud services, both for computing and storage. Initial content and derivatives are stored and processed in the cloud, often in a distributed manner. Furthermore, authentication, authorization and accounting are aspects that OpenMinTeD needs to look into as an integral part of the platform. It is important that we do not start fresh, but build up the synergies with existing European e-Infrastructures and take advantage of shared resources. This will ensure an increased uptake for the OpenMinTeD services as they will use current investments.

Who to collaborate with?	Why?	Expected Outcome
EUDAT	EUDAT is a major European e-Infrastructure for data management and preservation services. Provided the stability of EUDAT's services, OpenMinTeD will promote them to its constituency.	Use of EUDAT's services: ▪ B2SHARE repositories to deposit annotated results. ▪ B2SAFE to copy content from different storage locations. Share best practices for Open Science (legal). Promote common protocols & formats for access to content. Share technology on workflows & annotations.
EGI	EGI's federated cloud approach is a model to be potentially used in the OpenMinTeD architecture. GRNET is the liaison to EGI's open stack.	Use of shared cloud resources. Promotion of Open Science via EGI's communities. Share know-how on accounting services.
GEANT	Apart from network technology Geant provides AAI mechanisms that may be used in sharing of resources.	Adoption and use of EduGain and other AAI protocols.

Who to collaborate with?	Why?	Expected Outcome
Indigo Data Cloud	This is a H2020 project aiming to develop a data and computing platform targeting scientific communities, deployable on multiple hardware and provisioned over hybrid (private or public) e-infrastructures.	Use of shared cloud resources.
e-IRG	The e-Infrastructure reflection group is bringing together national and European e-Infrastructures and ESFRIs. OpenMinTeD services should be part of this ecosystem.	Outreach to e-Infrastructure and ESFRI communities. Develop KPIs, costs, services for the European service catalogue.
Europeana	Europeana's LoCloud project supports small and medium-sized institutions in making their content and metadata available to Europeana, by exploring the potential of cloud computing technologies.	Knowledge transfer, use of common resources.
OADA	The Open Ag Data Alliance is a US-based initiative that works on enhancing farmers' data interoperability and security through open REST APIs	Explore opportunities of synergies (e.g. adoption of TDM services by OADA) and use as additional data source.

3.1.4 SMEs, Industrial players

TDM is currently changing the way scientists conduct their business. It has influenced research processes and broadened up the area with new players and has attracted new businesses. With OpenMinTeD in place these companies can have homogenized access to scientific literature, and can discover tools or services to produce innovative products that they can make discoverable through the registry.

The table below lists a few examples of SMEs that the OpenMinTeD consortium has connections to. The aim is to broaden up this list with the second year open calls for SMEs and research teams to use the OpenMinTeD platform (guidelines, APIs, services).

Who to collaborate with?	Why?	Expected Outcome
LT-Innovate	Language Technology Industry Association. Provides outreach potential through its registry of SMEs/vendors.	Networking and outreach to LT and NLP related SMEs.
Linknovate	A startup that text mines scientific publications to produce research analytics (topic modeling)	Registration in the platform. Use of common APIs to access primary content and to provide services.
UberResearch	A research analytics startup (in the Digital Science group) that provides services to funders. It does so through topic modeling and related NLP mechanisms.	Registration in the platform. Use of common APIs to access primary content and to access/provide services.
ResearchFish	A UK based company that provides a research outcomes collection and evaluation service for Funders, Researchers and Research Institutes.	Registration in the platform. Use of common APIs to access primary content and to access/provide services.
Ontotext	Ontotext combines a semantic graph database with text mining to produce services	Use of common APIs to access primary content and to access/provide services.
ContentMine	A startup that produces software and training resources to help researchers, educators, citizens and others to apply TDM on open literature.	Registration in the platform. Use of common APIs to access primary content and to access/provide services.
Linguamatics	A TDM company focused on life sciences. Has developed the I2E text NLP/mining platform	Use of common APIs to access primary content and to provide services.

3.1.5 Funders and Ministries

This group of stakeholders is essential as they are key in changing the Open Science, and consequently TDM related, policies and influence how these are implemented. OpenMinTeD will collaborate with them in the following ways, with the aim to help overcome the technical and legal limitations of machine access to research publications:

- raise awareness via the platform, services/content and applications to be used by researchers (success stories);
- provide research analytics services combined with OA content from OpenAIRE, CORE and other OA publishers.

Who to collaborate with?	Why?	Expected Outcome
EC and ERC	The EC and ERC officials have already understood the importance of TDM and how this is part of Open Science (Commisioner Moedas "Open Innovation, Open Sciene, Open to the World").	Collaboration on how to achieve better TDM policies, improved Open Science, improved innovation. Use of OpenMinTeD services for research analytics.
National funding authorities	As they are developing Open Access and data management policies (e.g. all publications and data out of a funded project to have open licenses allowing TDM), national funders are getting interested in Open Science initiatives and e-Infrastructures. Will combine the outreach efforts via OpenAIRE.	Serious impact on the flow of open content and data.
Wellcome Trust	A UK funder who collaborates with PMC Europe for TDM services	Endorsement of the platform and guidelines. Potential use of more TDM services via the OpenMinTeD platform.
Science Europe	An organization with outreach to many funding organizations.	Network and outreach. Recommendations for Open Science TDM related policies.

3.1.6 Legal Experts and Policy Makers

TDM legal aspects related to content and service provision are a very challenging aspect in OpenMinTeD. Even though the technical infrastructure will be developed based primarily on OA content, copyright implications or license impositions and restrictions make TDM processes non

trivial. Furthermore, legal aspects related to service provision will need special focus for a proper uptake of the OpenMinTeD platform by researchers and SME's.

OpenMinTeD will liaise with a variety of organizations, mainly to intensify awareness and highlight good practice examples and use them for demonstrating the effects of data sharing.

Who to collaborate with?	Why?	Expected Outcome
FutureTDM	The *twin* H2020 project deals with the promotion and homogenization of TDM content licenses. The common project partners (LIBER, UVA, ARC) will ensure that complementary dissemination plans raise awareness to the proper stakeholder groups.	Push for TDM exceptions on copyright and database right laws. Develop a common knowledge base on TDM related topics. Success stories. Raise awareness of legal issues to various stakeholders. Promotion of the OpenMinTeD guidelines and platform.
Creative Commons	Provides licenses for research related material.	Build on CC's know-how, promote/use CC licenses. Develop success stories.
NEXA/Communia	COMMUNIA advocates for policies that expand the public domain and increase access to and reuse of culture and knowledge. It essentially advocates for improvements to the EU copyright framework.	Liaise to raise awareness on TDM exceptions for research related content.
EC IPR Helpdesk	EC's Helpdesk on how to manage IP and IPRs. Provide first-line support to beneficiaries of EU funded research projects.	Share knowledge base, FAQs and other support/training material.
Open Knowledge	A worldwide non-profit network of people passionate about openness, using advocacy, technology and training to unlock information and enable people to work with it to create and share knowledge. TDM is on their radar, together with OA.	Use OK channels to raise awareness on TDM policy issues. Promote OpenMinTeD platform through the OK channels (mailing lists, events, etc.)
IFLA, LIBER, ARL	These are the main bodies representing the interests of library (research or other) and information services and their users.	Raise awareness on TDM licenses and costs. Support/train librarians to promote OpenMinTeD to their researchers.

Who to collaborate with?	Why?	Expected Outcome
OpenAIRE	Outreach to repository managers.	Promote the correct licenses for repository content.
META-SHARE GateCloud (previously AnnoMarket)	Service provision legal issues have already been considered in Meta-Share and AnnoMarket (now superceded by GateCloud). OpenMinTeD will support open services provided by text mining researchers who are already aware of these issues.	Provide incentives for text miners and language researchers to engage in the infrastructure. Make conditions clear for SME's who act as service providers.
WIPO	A global forum for intellectual property services, policy, information and cooperation.	Follow and learn from the forum's activities. Raise awareness to Europe's IPR and licensing ecosystem.
RDA/CODATAT IG on legal interoperability	This group consists of international experts on IPR and licenses of research related data.	Discussions and visibility of TDM issues on the global scene. Adoption of IG's principles.

3.1.7 Research Communities

Apart from the OpenMinTeD project community partners, there is a large number of European research communities that are keen on using TDM services on scientific literature or other related content. Besides extracting knowledge for their domain, they can use TDM services to enrich metadata describing research data. The following table presents a set currently known to the consortium, but OpenMinted will open up to more research communities as opportunities arise.

Who to collaborate with?	Why?	Expected Outcome
ELIXIR	A European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society.	Outreach to the Life Sciences Community
Instruct	A pan-European research infrastructure in structural biology, making high-end technologies and methods available to users.	Outreach to the Life Sciences Community
Phenomenal	An e-infrastructure for clinical metabolomics data	Embed OpenMinTeD interoperability framework in the newly formed e-Infrastructure.
INFRAFRONTIER	The INFRAFRONTIER Research Infrastructure provides access to mouse models, data, and scientific platforms and services to study the functional role of the genome in human health and disease.	Outreach to the Life Sciences Community
CLARIN, DARIAH	Digital research infrastructures for the Humanities and Social Sciences to access and analyze data as well as depositing data and research results	Outreach Digital Humanties communities
AGINFRA	AGINFRA is a global hub of agri-food research stakeholders, such as researchers, initiatives, projects and organizations.	Outreach to a large agri-food research community that could benefit from TDM services and applications. Possible adaptation and adoption of OpenMinTeD's outcomes in various instances.

Who to collaborate with?	Why?	Expected Outcome
Europeana	A portal providing access to millions of books, paintings, films, museum objects and archival records that have been digitised throughout Europe.	Outreach to cultural heritage organizations and national libraries.
EPOS ENVRI	Outreach to environmental researchers scientific domain.	Link to related data infrastructures for the provision of TDM services.
Farr Institute	UK central Health Informatics group, with a growing interest in TDM	Source of clinical date related requirements.
Automation and Systematc Reviews Group	A UK and US based group looking at automation of systematic reviews of the medical literature. Outreach to this community.	Potential source of further requirements, and possible end users.
Research Data Alliance (RDA) - IGAD	The Agricultural Data Interest Group (IGAD) of RDA attracts a number of key stakeholders in agri-food research at a global level	Promote OpenMinTeD's expected outcomes to a global domain-specific research community and explore opportunities for adoption.

3.1.8 *"Standardization" bodies and fora*

Although at present there are many converging developments, different NLP infrastructures and tools often implement only some of these standards, which hampers interoperability. Therefore, given the existence of this variety of (standard) linguistic models, it is necessary to establish interoperability between their vocabularies in a principled way, in order to enable text mining tools to be brought together within the OpenMinTeD platform.

Who to collaborate with?	Why?	Expected Outcome
WC3 groups Ontology-Lexica Community Open Annotation Community	Inform about standardized programmatic interfaces and access rules.	Identify and pursue opportunities for the introduction of new or enhancement of existing standards.
OpenAIRE, Jisc/Rioxx, CASRAI, COAR, DataCite	These are the primary initiatives that come up with guidelines for metadata and content access.	Use and promotion of common protocols and formats for metadata and resource description.
RDA	RDA seems to be an excellent forum to discuss legal and technical issues related to the access of content, resource description (including workflows and annotations).	Promotion of OpenMinTeD guidelines for data and service sharing. Possible formation of an interest or working group related to OpenMinTeD's working groups.
NISO	The NISO access and License Indicators are used by publishers and has a "*Free to Read*" attribute, which may leave TDM in a limbo state.	Promotion of OpenMinTeD guidelines for TDM uptake.
CLARIN META-SHARE	These are two of the primary initiatives that have dealt with metadata descriptions and especially with schemas for language resources.	Build on existing data models. Use/promote common protocols and formats for resources.
Force11	A community led initiative to improve future research communication and e-Scholarship	A forum to help raise awareness on TDM issues for legal and technical interoperability.

• • •

3.1.9 Linked Open Data Initiatives and systems

A key text mining interoperability challenge is that linguistic descriptions come from existing language resources and tools, such as thesauri, lexical databases and linguistic annotation tools, and often content of interest. Linked Open Data (LOD) mechanisms and protocols are already in use by the text mining research community, and we need to ensure that there is no duplication of work through collaboration with the following initiatives:

Who to collaborate with?	Why?	Expected Outcome
Linguistic Linked Open Data	Provides information about the current status of the growing cloud of linguistic linked open data.	Identify and pursue opportunities for the introduction of new or enhancement of existing standards. Use LOD datasets and registry. Promote OpenMinTeD platform.
LOD/NIF	NLP Interchange Format (NIF) is an RDF/OWL-based format that allows to combine and chain several NLP tools in a flexible, light-weight way. In addition W3C has issued a preliminary set of guidelines on NIF-based web services.	Identify and pursue opportunities for the introduction of new or enhancement of existing standards. Use/share of LOD-related protocols. Promote OpenMinTeD platform.
PSI related data	Europe's open data portals ▪ www.europeandataportal.eu/en ▪ www.open-data.europa.eu ▪ www.publicdata.eu are getting richer in content and contain TDM research related corpora.	Access to textual, unstructured content. Promotion of common guidelines and APIs.
DBPedia	Contains structured information from Wikipedia and makes this information available on the Web through LOD mechanisms.	Access to primary content to combine with research publications and use in OpenMinTeD's use cases.
Open Knowledge	The Open Linguistics Working Group of the Open Knowledge Foundation works towards a linked open data cloud of linguistic resources, which applies the linked data paradigm to linguistic knowledge.	Promotion and use of common formats and protocols, guidelines.

Who to collaborate with?	Why?	Expected Outcome
Open Data Institute (ODI)	ODI is the leader in supporting open data activities and promoting data-based innovation.	Explore opportunities for collaboration and adoption of project's outcomes, establish connection with ODI's data startups who may identify innovative applications based on the project's outcomes.

3.1.10 International initiatives

OpenMinTeD is the European initiative for text mining, but there are similar or complementary initiatives around the world. As some of these initiatives tackle the same problems (interoperability of content access/retrieval, service provision over the cloud, AAI to name a few) the OpenMinTeD consortium will establish liaison activities with them in order to i) avoid duplication of work, and ii) aim for globally interoperable infrastructures.

Who to collaborate with?	Why?	Expected Outcome
LAPPS Grid (US)	The LAPPS Grid provides facilities to select from hundreds of NLP tools to create workflows, composite services, and applications, and to evaluate, reproduce, and share them with others	Common interoperability framework components.
Language Grid (Japan)	Language Grid is an online multilingual service platform which enables easy registration and sharing of language services such as online dictionaries, bilingual corpora, and machine translators.	Common interoperability framework components.
Alveo (Australia)	Alveo provides an infrastructure for accessing human communication data sets and to tools and services for processing and annotating that data	Common interoperability framework components.
DeepDive (US) PaleoDeepDive (US)	A data management framework that enables extraction, integration, and prediction problems in a single system, allowing users to rapidly construct sophisticated end-to-end data pipelines. PaleoDeepDive is an instantiation in the Univ. of Wisconsin, Madison to serve the paleontology research community.	Identify framework or infrastructure architecture components. Investigate synergies on interoperability aspects. Look into UW Madison licensing contracts with publishers.

Who to collaborate with?	Why?	Expected Outcome
HATHI TRUST (US)	The HathiTrust repository with its ~500TB of digitized, restricted data is a latent goldmine for text mining analysis, analysis of large-scale corpi through computational tools, and time-based analysis. Prevailing philosophy is that computation moves to data.	Investigate metadata descriptions used and overall big data approach. Promote/use common protocols and formats.
Domeo (US)	Domeo offers an extensible web application enabling users to visually and efficiently create and share ontology-based stand-off annotation. It supports manual, fully automated, and semi-automated annotation, individual or community-based, with appropriate access control and provenance recording.	Inspiration and synergies in the specification and, potentially, development of the OpenMinTeD annotation editing environment.
ARL/SHARE (US)	US aggregator of scientific publications.	Promotion of common protocols and formats. Consumer of OpenMinTeD services and tools.
La Referencia (Latin America)	LA aggregator of scientific publications.	Promotion of common protocols and formats. Provider (broker) of content and consumer of OpenMinTeD services and tools.
GODAN (Global)	The Global Open Data for Agriculture and Nutrition (GODAN) is a global initiative that aims to improve sharing of open data to make information about agriculture and nutrition available, accessible and usable.	Adoption of OpenMinTeD services and tools and application over existing open data repositories aiming to enhance data availability, discoverability and accessibility.
The Global Food Safety Partnership (GFSP)	GFSP is led by the World Bank and engages business and industry, governments, regulatory bodies, international development organizations, and civil society, working on food safety training & improving skills, knowledge and resources in the sector.	Exploration of collaboration opportunities through participation in events like the GFSP Annual Meetings and GFSI Conferences. Adoption of the OpenMinTeD outcomes on the GFSP data

4. *Approach*

The table below illustrates the approach methodology for the various stakeholders.

When?	Stakeholder group	How to approach this stakeholder?
High Priority	Publishers, scholarly societies and repositories	Direct communication via Frontiers and LIBER and planned workshops. Use OpenAIRE NOADs and its services to extend agreements to Use CORE UK outreach.
	Text mining and language researchers	Direct communication via engagement in the OpenMinTeD WP 5.2 as external experts on the interoperability specification and related workshops, as well as via project partners where there is an overlap between OpenMinTeD consortium members and research communities. Where this is not the case, direct approach to the communities.
	Cloud and Data infrastructure	Direct communication via project partners (ARC, GRNET, INRA). Liaise via EC concertation and related meetings. RDA working groups.
	Legal Experts and Policy Makers	Join efforts with FutureTDM for broad dissemination.
Medium Term	Linked Open Data Initiatives and systems	Participate in events organized by the initiatives ensuring exploration of opportunities for collaboration and adoption of project's outcomes.
	Research Communities	Via e-IRG and RDA fora. Publication of OpenMinTeD services in the upcoming e-Infra service catalogue.
	Researchers	Use OpenMinTeD research communities to pass on the message. Join forces with FutureTDM to raise awareness for TDM and OpenMinTeD services.
	SMEs and Industrial players	Use the GateCloud outreach. Direct approach to SMEs via EC's EINFRA-22 call.
	Funders and Ministries	Promote via OpenAIRE NOADs who have access to local Use EC's National Reference Points for OA.