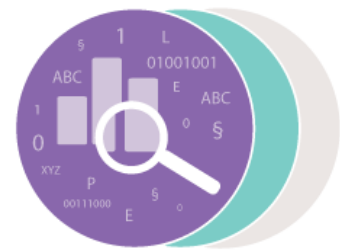




# FutureTDM

Explore . Analyse . Improve



REDUCING BARRIERS AND INCREASING UPTAKE OF TEXT AND DATA MINING FOR RESEARCH ENVIRONMENTS USING A COLLABORATIVE KNOWLEDGE AND OPEN INFORMATION APPROACH

## Deliverable D1.4

### Data Management Plan

(updated at Month 9)

## Project

Acronym: **FutureTDM**

Title: Reducing Barriers and Increasing Uptake of Text and Data Mining for Research Environments using a Collaborative Knowledge and Open Information Approach

Coordinator: SYNNO GmbH

Reference: 665940

Type: Collaborative project

Programme: HORIZON 2020

Theme: GARRI-3-2014 - Scientific Information in the Digital Age: Text and Data Mining (TDM)

Start: 01. September, 2015

Duration: 24 months

Website: <http://www.futuretdm.eu/>

E-Mail: [office@futuretdm.eu](mailto:office@futuretdm.eu)

Consortium: **SYNNO GmbH**, Research & Development Department, Austria, (SYNNO)  
**Stichting LIBER**, The Netherlands, (LIBER)  
**Open Knowledge**, UK, (OK/CM)  
**Radboud University**, Centre for Language Studies The Netherlands, (RU)  
**The British Library Board**, UK, (BL)  
**Universiteit van Amsterdam**, Inst. for Information Law, The Netherlands, (UVA)  
**Athena Research and Innovation Centre in Information, Communication and Knowledge Technologies**, Inst. for Language and Speech Processing, Greece, (ARC)  
**Ubiquity Press Limited**, UK, (UP)  
**Fundacja Projekt: Polska**, Poland, (FPP)

## Deliverable

Number:	<b>D1.4</b>
Title:	<b>Data Management Plan</b>
Lead beneficiary:	<b>SYNYO</b>
Work package:	WP1 Manage: Project management and collaboration
Dissemination level:	Public (PU)
Nature:	Report (RE)
Due date:	29.02.2016
Submission date:	29.02.2016
Author:	<b>Alessio Bertone, SYNYO</b>
Contributors:	<b>Helen Frew, LIBER</b> <b>Melanie Imming, LIBER</b> <b>Freyja van den Boom, OKCM</b> <b>Stelios Piperidis, ARC</b>
Review:	<b>Stelios Piperidis, ARC</b>

**Acknowledgement:** This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 665940.

**Disclaimer:** The content of this publication is the sole responsibility of the authors, and in no way represents the view of the European Commission or its services.



# Table of Contents

- 1 Introduction..... 6
- 2 Dataset management..... 7
- 3 Data sharing..... 9
- 4 Archiving and preservation ..... 11
- 5 Ethics ..... 12
- 6 References..... 13
- 7 Annex..... 14

## List of Tables

Table 1: Relevant aspects in dataset management..... 7

Table 2: Access procedure and access rights followed by FutureTDM..... 9

Table 3: Dissemination channels..... 10

Table 4: Storage and preservation in FutureTDM..... 11

## 1 INTRODUCTION

This deliverable describes the Data Management Plan (DMP) for the FutureTDM project. The aim of the DMP is to provide an analysis of the main elements of the data management policy that will be used throughout the project, with regard to all the datasets that will be generated and used. The DMP follows the guidelines suggested for Horizon 2020 calls<sup>1</sup> and will evolve during the lifespan of the project. Moreover, an ethical approach will be adopted and maintained throughout the fieldwork process, following the directives described in deliverable 2.1.

---

<sup>1</sup>[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

## 2 DATASET MANAGEMENT

This section encompasses the current status within the consortium about the data that will be produced. It will be coherently updated as the project progresses to always reflect the current status. In particular, Table 1 outlines the most relevant aspects that the FutureTDM project will take into account.

**Table 1: Relevant aspects in dataset management**

Data set reference and name		<i>Structured interviews</i>
Data set description	General description	The data gathered concerns the answers given by TDM practitioners on questions related to text and data mining practices and issues. In total the 30 interviews will be recorded to mp3 and transcribed into an excel file. The data will be made available anonymized to the project partners for use within the FutureTDM project
	Provenance	Face to face interviews
	Nature	Mixed method research. The questionnaire contains questions on the participants' professional opinion and experiences with TDM. No additional or sensitive personal data is being collected
	Scale	30 stakeholders
	Beneficiary	The data is collected for internal use of the FutureTDM project. However the results will feed back into the FutureTDM project through visualizations and reports, to derive insight and provide best practices which will be made publicly available
Standards and metadata		The interviews will be recorded (mp3), transcribed together with additional notes in Excel and reported in form of a document. In case the primary interview data (actual mp3 files) are to be persistently stored for future use by other researches, they would be appropriately described using a compatible schema (Dublin Core)
Data collection procedure		<ul style="list-style-type: none"> <li>Stakeholder collection/ KC participant list: names, contact details and organisations as part of the stakeholder identification will be collected and (after getting their consent) made available in the online stakeholder map</li> <li>Bibliographic references: references will be collected and stored (e.g. in bibTex format), but being public domain data, no ethical issue arises</li> </ul>

	<ul style="list-style-type: none"> <li>• Structured Interviews (WPs 4 and 5): participants are chosen from the stakeholder collection and, based on their voluntary informed consent, asked to answer a set of pre-determined open and closed question in person or through videocall. With the participants consent (using the form provided in the Annex), the interviews will be recorded and transcribed for internal use only (adopting the standards above mentioned). The data will be anonymized for further use for the purpose of FutureTDM</li> <li>• Mainly public data will be used throughout the FutureTDM project. In case industrial projects are used, the owner of the Intellectual Property Rights will be approached and will have to approve that the data can be used for the project. Furthermore, the data will be aggregated and anonymized for ensuring that personal and or confidential data are not violated</li> <li>• Interviews (WPs 2 and 7): Video and photo images will be made available on the project website under Creative Commons BY Attribution v. 4.0 License and may be played at the larger multi-stakeholder workshops/symposium. The participants will be asked to sign a “consent form”, where the participants confirm that all portraits and images are made with the explicit authorization of the participant. The participants also confirm that the FutureTDM project can use the videos and images for the FutureTDM projects</li> </ul>
--	--



### 3 DATA SHARING

Table 2 outlines access procedures and rights in relation to the data gathered throughout the FutureTDM project

**Table 2: Access procedure and access rights followed by FutureTDM**

Access procedure	In accordance with Grant Agreement Article 25, data must be made available upon request, or in the context of checks, reviews, audits or investigations. If there are ongoing checks etc. the records must be retained until the end of these procedures
Access rights	<p>Project partners:</p> <ul style="list-style-type: none"> <li>• FutureTDM partners must give each other access — on a royalty-free basis — to data needed to implement their own tasks under the action, where is legally and practically possible</li> <li>• FutureTDM partners must give each other access – under fair and reasonable conditions (Article 25.3) – for exploiting their own results to data, where is legally and practically possible</li> <li>• Unless otherwise agreed, requests for access may be made up to one year after the period set out in Article 3 (24 months)</li> </ul> <p>Affiliated entities:</p> <ul style="list-style-type: none"> <li>• Unless otherwise agreed, access must be given, under fair and reasonable conditions, and where is legally and practically possible</li> <li>• Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3 (24 months)</li> </ul>

Concerning the exploitation and the dissemination of results, each partner must take measures to ensure the exploitation of its results, up to four years after the period set out in Article 3 (24months) and to guarantee the access and visibility of the results (according to Article 29 of the Grant Agreement). To this aim different dissemination channels are adopted, improved and maintained also after the project lifecycle (for more detailed information see D7.2, Communication and exploitation plan). They are shown in Table 3 along with a short description about their use and the policies adopted. The content presented in the table will be coherently updated as the project progresses.

**Table 3: Dissemination channels**

<b>Dissemination channels</b>	<b>Usage</b>	<b>Policy</b>
Project website	Reference point of project visibility until the Open Information Hub goes online	CC-BY
Newsletter	Provide regular updates on the project activities and redirect to the website, where more information on the project is available	CC-BY
Fact sheets	Support the work of the project and encourage feedback e.g. at events	CC-BY
Knowledge Cafés (KC)	Informal opportunity for stakeholders to find out about TDM, the FutureTDM project and its goals, and to provide the project with feedback	Chatham house rule <sup>2</sup>
KC flyers	Explain knowledge cafés and asking for input	CC-BY
Social media (e.g. Twitter)	Publicise the project several time a day and support the diffusion of TDM related news	CC-BY
Publications	Project related articles in TDM field	Open Access
Blog	A place where stakeholders can find latest updates on the project, useful info and exchange comments on TDM related topics	CC-BY
Templates	Ensure brand continuity	CC-BY
Project reports	Describe the results of the work packages	CC-BY
Video	Gain an insight into FutureTDM and involve stakeholder to improve TDM uptake in EU	Consent form + CC-BY
Survey (e.g. structured interviews)	Collect experts feedback and generate best practice case studies	Consent form + CC-BY for best practice
FutureTDM Platform	Contain blogs, events, printable materials etc. to discuss and promote TDM in Europe	CC-BY

<sup>2</sup> [https://en.wikipedia.org/wiki/Chatham\\_House\\_Rule](https://en.wikipedia.org/wiki/Chatham_House_Rule)

## 4 ARCHIVING AND PRESERVATION

Table 4 outlines the main management principles behind the archiving and preservation of the data collected through the project.

**Table 4: Storage and preservation in FutureTDM**

Inform and keep track	<ul style="list-style-type: none"> <li>• Data gathered in WP2 (interviews and workshops) and WP4 (best practices) as well as their metadata, will be compiled and deposited in OpenAIRE's Zenodo repository to ensure discoverability, accessibility, and intelligibility</li> <li>• In case of changes to this regard, each partner must immediately inform the coordinator (who in turn must inform the Funding Agency and other partner countries)</li> <li>• Records and documentation will kept up-to-date in content and format so they remain easily accessible and usable</li> </ul>
Retention	A period of four years (after the end of the project)
Type of documents retained	Project partners retain the original documents. Digital and digitalised documents are considered originals if they are authorised by the applicable national law

## 5 ETHICS

The project partners are to comply with the ethical principles as set out in the Grant Agreement (Article 34), which states that all activities must be carried out in compliance with:

- The ethical principles (including the highest standards of research integrity e.g. as set out in the European Code of Conduct for Research Integrity, and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct) and Commission recommendation (EC) No 251/2005 of 11 March 2005 on the European Charter for Researchers and on a Code of Conduct for the Recruitment of Researchers (OJ L 75, 22.03.2005, p. 67), the European Code of Conduct for Research Integrity of ALLEA (All European Academies) and ESF (European Science Foundation) of March 2011
- Applicable international, EU and national law.

Furthermore, activities raising ethical issues must comply with the “ethics requirements” set out in Annex 1 of the Grant Agreement.

### **Confidentiality**

Whenever not differently written, the FutureTDM partners must retain any data, documents or other material as confidential (“confidential information”) during the implementation for the project and for four years after the period set out in Article 3 (24 months). Further details on confidentiality can be found in Article 36 of the Grant Agreement.

## 6 REFERENCES

### Literature:

European Union Agency for Fundamental Rights (2014) *Handbook on European Data Protection Law*. Brussels.

### Websites:

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

## 7 ANNEX

As described in Deliverable 2.1, the participants to the survey will receive a form like the following one:

<b>Purpose of data collection:</b> <i>The following data will be collected for the EU research project FutureTDM funded by European Commission under the Horizon2020 – GARRI-3 program. FutureTDM partner will use this data for Work Package _____.</i>	
<b>FutureTDM Consortium Contact Point(s):</b>	
<b>Who has access to this information:</b> By signing the form you give your consent to process your data for the project. The <i>FutureTDM</i> consortium members will be the only persons that will have access to your personal information. The <i>FutureTDM</i> Consortium members, who see/access this information, will keep it confidential.	
<b>Withdrawal Information:</b> Your participation in the <i>FutureTDM</i> project is completely voluntary, and you can choose to stop participating at any time. If you decide to withdraw from the project, please contact the <i>FutureTDM</i> consortium contact points outlined above, and they will explain the best way for you to stop taking part. You should know that you may be withdrawn from the project for any of the following reasons:	
<ul style="list-style-type: none"> <li>• If you don't follow the project's ethical committee instructions.</li> <li>• If you don't attend the scheduled data collection sessions.</li> <li>• If the whole project is stopped, for reasons not known now.</li> </ul>	
<b>Voluntary Participant Data</b>	
<i>Name</i>	
<i>Status</i>	
<i>Email</i>	
<i>Telephone</i>	
<b>Chair of Selection Panel of Voluntary Participant</b>	
<i>Name</i>	
<i>Address</i>	
<i>Email</i>	
<i>Telephone</i>	
<b>Exercise details</b>	
<i>Exercise Plan Form</i>	(Id number)
<i>Actor Role Form</i>	(Id number)
<b>Applicable Laws/Directives</b>	
<i>Date</i>	(dd/mm/yy)
<i>Declaration</i>	I have read the terms outlined and understand them. I consent to the terms.  <i>Signature</i> .....