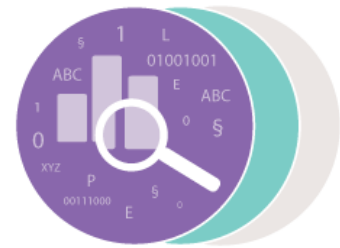




FutureTDM

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 D4.4

Scientific Review Paper

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

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:	D4.4
Title:	Scientific Review Paper
Lead beneficiary:	RU
Work package:	WP4: Fields of Application, Projects, Best Practices and Resources
Dissemination level:	Public (PU)
Nature:	Report (RE)
Due date:	30.04.2017
Submission date:	02.05.2017
Authors:	Maria Eskevich, RU Antal van den Bosch, RU
Contributors:	Stelios Piperidis, ARC Kanella Pouli, ARC Maria Gavriilidou, ARC Dimitris Galanis, ARC
Review:	Burcu Akinci, SYNYO

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 does not in any way represent the view of the European Commission or its services.

This report by FutureTDM Consortium members can be reused under the [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) license (<https://creativecommons.org/licenses/by/4.0/>).

Table of Contents

- 1 Summary..... 5
- 2 Work on the scientific paper 6
 - 2.1 Criteria for selection of potential publication venues..... 6
 - 2.2 Information on the submitted paper 8
- 3 Work on the FutureTDM knowledge base enrichment..... 9
 - 3.1 Collection of Projects..... 9
 - 3.2 Collection of Organisations 9
 - 3.3 Collection of Methods 10
 - 3.4 Collection of Tools 11
- 4 Conclusions 15
- 5 Annex..... 16

List of Tables

- Table 1: Total number of projects in the FutureTDM Knowledge Base Collection structured by economic sector 9
- Table 2: Total number of projects in the FutureTDM Knowledge Base Collection structured by country 10
- Table 3: FutureTDM text and data mining methods collection 11
- Table 4: Total list of tools in the FutureTDM Knowledge Base Collection 14

1 SUMMARY

This report provides information about the work on the scientific review paper that gives a critical review of the past, present and future of the state-of-the-art of the field of text and data mining (TDM). This publication aims to broaden perspective on the past, current, and future trends in TDM amongst audience, outlined via desktop research, consultation with the practitioners (Deliverables D.4.3. and D4.5.) and scientific communities. Report includes brief description of publication venue selection, text of the paper that can be further updated with the open accesses reviewers comments.

Additionally, the information about the update of FutureTDM knowledge base collection of relevant projects, organisations, methods and tools is provided including the current list of the collections in the Annex.

2 WORK ON THE SCIENTIFIC PAPER

2.1 Criteria for selection of potential publication venues

We have defined the two main criteria for the target publication venues:

- Diverse type of TDM applications already present in the journal
- The journal has to be Open Access, or at least to have the open access publication option

As the review of TDM work from the perspective of overall scientific development and insights on its visibility across different economic fields, that are formulated within the FutureTDM consortium, differ from the traditional type of publication content that is expected by journals, we took an initiative to contact a number of targeted journals in order to find a venue that would fit our criteria and to ensure that our content would be suitable. These were the journals to contact:

1. Data Science Journal

- Link: <http://datascience.codata.org>
- Aims and scope of the journal: The CODATA Data Science Journal is a peer-reviewed, open access, electronic journal, publishing papers on the management, dissemination, use and reuse of research data and databases across all research domains, including science, technology, the humanities and the arts. The scope of the journal includes descriptions of data systems, their implementations and their publication, applications, infrastructures, software, legal, reproducibility and transparency issues, the availability and usability of complex datasets, and with a particular focus on the principles, policies and practices for open data. All data is in scope, whether born digital or converted from other sources.
 - *Reviews* can cover topics such as current controversies, the current “state of the art” or the historical development of studies as well as issues of regional or temporal focus. Papers should critically engage with the relevant body of extant literature. Review articles should be no longer than 3,000 words in length.
- Open access policy: multidisciplinary Open Access journal

2. PLOSOne

- Link: <http://journals.plos.org/plosone/>
- Aims and scope of the journal: PLOS ONE features reports of original research from all disciplines within science and medicine. By not excluding research on the basis of subject area, PLOS ONE facilitates the discovery of connections between research whether within or between disciplines.

- We consider publishing systematic reviews only if the methods ensure the comprehensive and unbiased sampling of existing literature.
- Submissions describing methods, software, databases, or other tools. We consider submissions describing methods, software, databases, or other tools if they follow the appropriate reporting guidelines.
- Qualitative research. We consider publishing qualitative research only if it adheres to appropriate study design and reporting guidelines.
- Studies reporting negative results.
- Open access policy: multidisciplinary Open Access journal

3. DSH (Digital Scholarship for Humanities)

- Link: <https://academic.oup.com/dsh>
- Aims and scope of the journal: *DSH* or *Digital Scholarship in the Humanities* is an international, peer reviewed journal which publishes original contributions on all aspects of digital scholarship in the Humanities including, but not limited to, the field of what is currently called the Digital Humanities. Long and short papers report on theoretical, methodological, experimental, and applied research and include results of research projects, descriptions and evaluations of tools, techniques, and methodologies, and reports on work in progress. *DSH* also publishes reviews of books and resources.
- Open access policy: multidisciplinary Open Access journal

4. Science Communication (SAGE)

- Web-Link: <http://journals.sagepub.com/home/scx>
- Aims and scope of the journal: Science Communication is an international, interdisciplinary social science journal that examines such topics as the nature of scientific expertise as represented through communication and the processes or effects characterizing the communication of science in any context. Science is broadly defined to include environmental science, health science, and technology. Science Communication welcomes submissions of empirical research from authors in all relevant disciplines (including the social sciences, the humanities, and science itself). Both qualitative and quantitative research papers with a basis in theory are acceptable. Preference is given to articles that bridge the gap between theory and practice and that will be of interest across disciplines. In addition to peer-reviewed research, Science Communication publishes commentaries that analyse issues and trends in the field – whether scholarly, professional, or policy-related – and a periodic summary of new books in the field.
- Open access policy: OA option is available

5. IEEE Access

- Web-Link: <http://ieeaccess.ieee.org/learn-more-about-ieee-access/>

- Aims and scope of the journal: IEEE Access publishes articles that are of high interest to readers: original, technically correct, and clearly presented. The scope of this journal comprises all of IEEE's fields of interest, emphasizing applications-oriented and interdisciplinary articles. IEEE Access also accepts traditional technical articles, as well as reviews and surveys.
- Open access policy: Multidisciplinary Open Access Journal, fee \$1,750 per article

Eventually, we have chosen the “Data Science Journal” that is an open access multidisciplinary journal run by an SME publisher “Ubiquity Press”.

2.2 Information on the submitted paper

The details of the submission are the following:

- Authors: Maria Eskevich, Stelios Piperidis, Kanella Pouli, Maria Gavriilidou, Dimitris Galanis, Antal van den Bosch.
- Title: Text and Data Mining: Past, present, and future.
- Submission date: 01.05.2017.
- Text of the paper is in the Appendix

3 WORK ON THE FUTURETDM KNOWLEDGE BASE ENRICHMENT

3.1 Collection of Projects

FutureTDM collection of projects by the Deliverable submission date contains 140 items, 37 of which are already integrated into the platform. For 7 of the projects it is not possible to assign an economic sector, such as (+)Spaces , while the other 133 can be assigned to one or even a combination of sectors due to the multidisciplinary nature of the projects. Table 1 contains the overall statistics of the number of projects in the collection.

Sector or combination of Sectors that that the project is assigned to	Number of projects
Primary-Sector, Secondary-Sector, Tertiary-Sector, Quaternary-Sector	1
Primary-Sector, Quaternary-Sector	1
Secondary-Sector	8
Secondary-Sector, Tertiary-Sector	3
Secondary-Sector, Quaternary-Sector	38
Secondary-Sector, Tertiary-Sector, Quaternary-Sector	6
Tertiary-Sector	13
Tertiary-Sector, Quaternary-Sector	24
Quaternary-Sector	38
Quinary-Sector	1
TOTAL	133

Table 1: Total number of projects in the FutureTDM Knowledge Base Collection structured by economic sector

3.2 Collection of Organisations

FutureTDM collection of organisations by the Deliverable submission date contains 72 items, 59 of which are already integrated into the platform. 55 of these organisations have only one EU-country location, while three of them spread across more than one country within the EU, and 6 have locations outside EU in USA.

Company locations details	Country	Number
Company is based in one country	Austria	1
	Belgium	2
	Czech Republic	2
	Denmark	2
	Finland	1
	France	4
	Germany	10
	Greece	3
	Ireland	1
	Italy	3
	Norway	1
	Slovenia	1
	Spain	4
	Sweden	2
	Switzerland	4
The Netherlands	5	
UK	9	
Company has offices internationally	Belgium, The Netherlands, Germany, USA	1
	The Netherlands,	1
	Germany/USA	1
	Ireland/Greece	1
	Spain/UK	1
	UK/USA	2
	USA, many	2
Companies outside EU	Japan	1
	USA	7
TOTAL		72

Table 2: Total number of projects in the FutureTDM Knowledge Base Collection structured by country

3.3 Collection of Methods

FutureTDM collection of projects splits scientific methods into 29 items, and all of them are integrated into the platform, as they are used to specify the organisations focus.

TDM Methods			
Artificial Intelligence	Machine Learning	Statistics	Kernel Methods
Association rules	Machine Translation	Summarisation	Ensemble Learning
Classification	Multimedia Processing	Term/Concept Extraction	Dimensionality Reduction
Clustering	Natural Language Processing	Text Mining	Decision Trees
Data Mining	Negation and Modality Detection	Textual Entailment	Deep Learning
Graph Mining	Predictive Analytics	Artificial Neural Networks	
Information Extraction	Regression	Bayesian Inference	
Information Retrieval	Sentiment Analysis/Opinion Mining	Instance-Based Learning	

Table 3: FutureTDM text and data mining methods collection

3.4 Collection of Tools

FutureTDM collection of projects by the Deliverable submission date contains 72 items, 59 of which are already integrated into the platform. Table contains the complete list of tools collected so far.

Name	URL
ALVEO	http://alveo.edu.au/
Alvis	https://migale.jouy.inra.fr/redmine/projects/alvisnlp
Angoss Knowledge STUDIO	http://www.angoss.com/
AnnoMarket	https://annomarket.eu/
Apache cTAKES	http://ctakes.apache.org
Apache OpenNLP	http://opennlp.apache.org
Apache UIMA	https://uima.apache.org
Argo	http://argo.nactem.ac.uk

BioCatalogue	https://www.biocatalogue.org/
BiodiversityCatalogue	https://www.biodiversitycatalogue.org
Bluima	https://github.com/BlueBrain/bluima
CLARIN-DK	http://clarin.dk/
ClearTK	https://cleartk.github.io/cleartk
Clementine	http://datamining.togaware.com/survivor/Clementine.html
ContentMine	http://www.contentmine.org/
DKPro Core	https://dkpro.github.io/dkpro-core
ELKI	http://elki.dbs.ifi.lmu.de/
FICO Data Management	http://www.fico.com/
Galaxy	https://galaxyproject.org/
GATE Embedded	https://gate.ac.uk
Heart of Gold	http://heartofgold.dfki.de
IBM SPSS Predictive Analysis	http://www-03.ibm.com/software/products/en/spss-predictive-analytics-enterprise
JCoRe	http://julielab.github.io
Jpylyzer	http://jpylyzer.openpreservation.org/
KAF	http://kyoto-project.eu/xmlgroup.iit.cnr.it/kyoto/index2091.html?option=com_content&view=article&id=504&Itemid=173
Kepler	https://kepler-project.org/
KNIME	http://www.knime.org/knime
Language Grid	http://langrid.org/en/index.html
LAPPS Grid	http://www.lappsgrid.org/
Massive Online Analysis (MOA)	http://moa.cms.waikato.ac.nz/
Matchbox	http://matchbox.openpreservation.org/
Microsoft SQL Server	https://technet.microsoft.com/en-

Analysis Service (SSAS)	us/library/ms175609%28v=sql.90%29.aspx
Neural Designer	https://www.neuraldesigner.com/
NLTK	http://www.nltk.org
Open Calais	www.opencalais.com/
Oracle Data Miner GUI	http://www.oracle.com/technetwork/database/options/odm/dataminerworkflow-168677.html
Orange	http://orange.biolab.si/
Pagelyzer	http://pagelyzer.openpreservation.org/
Pegasus	https://pegasus.isi.edu
Pentaho	http://www.pentaho.com/
Pipeline Pilot	http://accelrys.com/products/collaborative-science/biovia-pipeline-pilot/
QT21	http://qt21.metashare.ilsp.gr/
Rapidminer Radoop	www.rapidminer.com
Rapidminer Server	www.rapidminer.com
Rapidminer Studio	www.rapidminer.com
SAP Hana	https://hana.sap.com
SAS Data Mining	http://www.sas.com
SPSS	http://www.ibm.com/analytics/us/en/technology/spss/
Taverna	http://www.taverna.org.uk/
Think Analytics	http://thinkanalytics.com/
Think Enterprise Data Miner	www.thinkanalytics.com/
Triana	http://www.trianacode.org
TTNWW	http://yago.meertens.knaw.nl/apache/TTNWW/
Weblicht	https://weblicht.sfs.uni-tuebingen.de/
WEKA	http://www.cs.waikato.ac.nz/~ml/weka
xcorrSound	http://xcorrSound.openpreservation.org/

VisTrails	http://www.vistrails.org
Trendminer	https://www.trendminer.com/
Opendatasoft	https://www.opendatasoft.com/
Meta	http://meta.com/
TensorFlow	https://www.tensorflow.org/about/
Epinium	http://epinium.com/
Exclusivi	http://exclusivi.com/hotels/
Eparkomat	http://www.eparkomat.com/
Data Scouts	https://datascouts.eu/
TopPlace	http://www.avuxi.com/products/topplace
Influency	https://influency.com/en/
Plazi TreatmentBank	http://plazi.org/api-tools/api/#Plazi_API
SafeToNet	https://safetonet.com/
Egas	https://demo.bmd-software.com/egas/
Wordfreak	http://wordfreak.sourceforge.net/
Theano	https://github.com/Theano/

Table 4: Total list of tools in the FutureTDM Knowledge Base Collection

4 CONCLUSIONS

This Deliverable contains the full text of submitted TDM review paper, together with the explanation on the publication venue selection.

We provide quantitative updates on the progress with filling in the four main categories of the Knowledge Base Collection, while the Deliverables 6.1, 6.2 and 6.3¹ contains the qualitative description of the technology behind the FutureTDM Platform.

¹ www.futuretdm.eu

5 ANNEX

In the following we present the scientific paper “**Text and Data Mining: Past, present, and future**” and the collections below described in the previous section. The list will be given in this order:

- Collection of Organisations
- Collection of Projects
- Collection of Tools

Collection of Organisations

Name	URL	Country	Zip	City	Street	Type	Data Sources	Sector	Application Field Sector I	Application Field Sector II	Application Field Sector III	Application Field Sector IV	Application Field Sector V
TheHub	http://thehub.com	The Netherlands	1088XK	Amsterdam	Science Park 402	Company	Business, Internet	Tertiary Sector, Quaternary Sector	N/A	N/A	N/A	Finance, Banking, Insurance, Research and Development (e.g. Tech)	N/A
Accenture	https://www.accenture.com	USA, many interns	N/A	N/A	N/A	Company	Business, Internet	N/A	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
ADAPT	http://adaptcentre.ie	Ireland	Dublin 2	Dublin	Trinity College, College Green	Research	All	Tertiary Sector, Quaternary Sector	N/A	N/A	Entertainment, Finance, Education, Research and Development (e.g. Tech)	N/A	N/A
Agri	http://agri.bio.gov.gr/agri-search/index.do	Greece	10776	Athens	Acharon Street 7	Company	Scientific	Primary Sector, Agriculture	N/A	N/A	N/A	N/A	N/A
AgriKnow	http://www.agriknow.com/agriknow/	Greece	152 35	Athens	Grimoussi 17	Company	Scientific	Primary Sector, Agriculture	N/A	N/A	N/A	N/A	N/A
Algorithmia	http://www.algorithmia.com/	The Netherlands	1075	Amsterdam	Orange Nassauboulevard 62-2	Company	All	Secondary Sector	Engineering	N/A	N/A	N/A	N/A
Alteos Limited	http://www.alteos.com/	Spain	37900	Salamanca	Carretera de Madrid 13	Company	Business, Personal	Secondary Sector, Tertiary Sector, Quaternary Sector	Bio-technology	Health and Social Care	Research and Development (e.g. Tech)	N/A	N/A
Artificial Intelligence Techniques Ltd	http://www.aitech.com/	UK	CB2 0RZ	Cambridge	Cambridge Business Park, Cowley Park	Company	All	Secondary Sector, Tertiary Sector	Manufacturing	Finance, Banking, Insurance, Health and Social Care, Public Administration	N/A	N/A	N/A
Autonomy	http://www.autonomy.com/	Spain/UK	N/A	N/A	N/A	Company	Business, Scientific	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A	N/A
AVUKI	http://www.avuki.com/	USA	N/A	Columbus, Ohio	N/A	Company	Business, Scientific	Secondary Sector, Quaternary Sector	Bio-technology, Manufacturing	Research and Development (e.g. Tech)	N/A	N/A	N/A
Bailela	http://www.bailela.com/	Austria	1040	Vienna	Schleichthausgasse 7	Company	Business, Public Sector	Secondary Sector, Tertiary Sector, Quaternary Sector	Bio-technology	Finance, Banking, Insurance, Education	N/A	N/A	N/A
BSI Ltd	http://www.bsi.com/	Denmark	2300	Copenhagen	Njalsgade 136, building 27	Research	Scientific	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A	N/A
Centre for Language Technology (CST)	http://csl.uu.se/csl/	Sweden	701 86	Uppsala	Ångström Laboratory, Ångström Laboratory	Research	Scientific	Secondary Sector, Quaternary Sector	Bio-technology	N/A	N/A	Research and Development (e.g. Tech)	N/A
ClearPath Clearance Center, Inc.	http://www.clearpath.com/	Denmark	2300	Copenhagen	Njalsgade 136, building 27	Research	Business, Scientific	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A	N/A
CrossRef	http://www.crossref.org/	UK/USA	OX1 1BF	Oxford	New Road	Research	Business, Scientific	N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Data Mind	http://www.datamind.co.uk/	Czech Republic	170 00	Prague	U Příhozu 22 / 466	Company	All	Tertiary Sector, N/A	N/A	N/A	Health and Social Care	Public Administration, Retail/Trade	N/A
DeepMind	http://www.deepmind.com/	UK/USA	N/A	N/A	N/A	Company	Business	N/A	N/A	N/A	N/A	N/A	N/A
Deutsche Forschungszentrum für Künstliche Intelligenz gmbh (DFKI)	http://www.dfki.de/	Germany	D-66113	Saarbrücken	Saarbrücken	Research	Scientific	Quaternary Sector	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
Dexar	http://www.dexar.com/	N/A	N/A	N/A	N/A	Company	All	N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
Dutch-Flemish Human Language Technology Agency ePronomat	http://www.epronomat.com/	The Netherlands, Belgium	N/A	N/A	N/A	Research	Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Edmentum	http://www.edmentum.com/	Czech Republic	N/A	N/A	N/A	Company	Business	N/A	N/A	N/A	N/A	N/A	N/A
Exclusiv Ltd.	http://www.exclusiv.com/	Spain	N/A	N/A	N/A	Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Expert Systems S.P.A.	http://www.expert-systems.com/	Ireland/Greece	N/A	N/A	N/A	Company	Business	N/A	N/A	N/A	N/A	N/A	N/A
Fondazione Bruno Kessler (FBK)	http://www.fbk.eu/	Italy	38122	Bozano	Via Santa Croce 77	Research	Scientific	Quaternary Sector	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
German National Library of Economics (ZBW)	http://www.zbw.eu/	Germany	68159	Mannheim	Quadrat 82, 1	Research	Scientific	N/A	N/A	N/A	N/A	N/A	N/A
GIS5: Leibniz-Institut für die Sozialwissenschaften	http://www.gis5.org/	Germany	68159	Mannheim	Quadrat 82, 1	Research	Personal, Scientific	Tertiary Sector, N/A	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
Helsinki University of Technology	http://www.aalto.fi/en/about/history/htk/	Finland	2150	Helsinki	Esopoo	Research	Scientific	Quaternary Sector	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
Institut national de recherche en informatique et en automatique (INRIA)	http://www.inria.fr/	France	78153	Roquecournot	BP 105	Research	Scientific	Quaternary Sector	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
Institut Universitari de Lingüística Aplicada (IULA)	http://www.iula.upc.edu/	Spain	8018	Barcelona	Universitat Pompeu Fabra Campus del	Research	Scientific	N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Institute for Language and Speech Processing (ILSP / "Athens" R.C.)	http://www.ilsp.gr/	Greece	151 25	Athens	Marousi, Artemidos 6	Research	Scientific	Quaternary Sector	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
Italian National Research Council (CNR)	http://www.cnr.it/	Italy	36124	Pisa	Italian National Research Council/Pisa	Research	Scientific	N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
JISC: Leibniz-Institut für die Sozialwissenschaften	http://www.jisc.ac.uk/	UK	852 6BA	Bristol	One Castlepark, Tower Hill	Research	Scientific	Quaternary Sector	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
Karlsruhe Institute of Technology	http://www.kit.edu/	Germany	76131	Karlsruhe	Kaiserstraße 12	Research	Scientific	Quaternary Sector	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
Komatsu	http://www.komatsu.com/en/aboutus/	N/A	N/A	N/A	N/A	Company	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Kudoss	http://www.kudoss.nl/	The Netherlands	N/A	N/A	N/A	Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Linguamatics	http://www.linguamatics.com/	UK	CB8 0WG	Cambridge	Milton Road	Company	Personal, Scientific	Secondary Sector, Tertiary Sector, Quaternary Sector	Bio-technology	Health and Social Care	Research and Development (e.g. Tech)	N/A	N/A
Ludwig-Maximilians-Universität	http://www.lmu-muenchen.de/index.html	Germany	80339	Munich	Geschwister Scholl Platz 1	Research	Scientific	N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
Max Planck Computing and Data Facility	http://www.rig.mpg.de/	Germany	85748	Garching	Gießereiachstraße 2	Research	Scientific	N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Mata	http://www.mata.com/	USA	N/A	N/A	N/A	Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
National Centre for Text Mining (NuCTM)	http://www.nuclm.ac.uk/	UK	M1 7DN	Manchester	Princess Street 111	Company	Business, Internet, Personal, Secondary Sector, Tertiary Sector, Quaternary Sector	Bio-technology	Health and Social Care	Education, Research and Development (e.g. Tech)	N/A	N/A	N/A
OpenDataSoft	http://www.opendatasoft.com/	USA/International	N/A	N/A	N/A	Company	Business, Internet	N/A	N/A	N/A	N/A	N/A	N/A
Paperkit	http://www.paperkit.net/	Switzerland	N/A	N/A	N/A	Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Piati	http://www.piati.com/	Norway	855	Oslo	Sognsvien 70A	Company	Business, Internet, Tertiary Sector, Quaternary Sector	N/A	N/A	N/A	Health and Social Care	N/A	N/A
Purple	http://www.purple.com/	Belgium	44227	Dortmund	Stoekumer Str. 475	Company	Business, Internet, Secondary Sector, Tertiary Sector	Bio-technology, Manufacturing	Entertainment, Finance, Banking, Insurance, IT Services, Retail/Trade	N/A	N/A	N/A	N/A
Python predictions	http://www.pythonpredictions.com/	Sweden	114 28	Stockholm	Brinellvägen 8	Company	Business, Scientific	N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
RapidMiner	http://www.rapidminer.com/	Germany	69190	Waldbrunn	Hasso Plattner Ring 7	Company	All	Primary Sector, Natural Resources Mining, Manufacturing	Engineering, Manufacturing	Entertainment, Finance, Research and Development (e.g. Tech)	N/A	N/A	N/A
SAP	http://www.sap.com/	USA	N/A	Washington, DC	N/A	Research	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Schubert Publishing and Academic Resources Coalition (SPARC)	http://www.sparc.org/	The Netherlands	10980X	Amsterdam	Science Park 402	Company	Business, Internet, Tertiary Sector, Quaternary Sector	N/A	N/A	N/A	Health and Social Care, R&D	Research and Development (e.g. Tech)	N/A
Slovenian Social Science Data Archives	http://www.ssdas.si/	Slovenia	1000	Ljubljana	Kongresni trg 12	Research	Personal, Scientific	Tertiary Sector, Quaternary Sector	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
Sponio	http://www.sponio.com/	USA	N/A	N/A	N/A	Company	Business, Scientific	Secondary Sector, Quaternary Sector	Bio-technology	N/A	Education, Research and Development (e.g. Tech)	N/A	N/A
Språkbanken, The Swedish language bank	http://www.sprakbanken.se/	Sweden	N/A	N/A	N/A	Research	Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Squares on Blue	http://www.squaresonblue.com/	France	78000	Versailles	Rue Saint Honoré 13	Company	All	Tertiary Sector	N/A	N/A	Finance, Banking, Insurance, Health and Social Care, Public Administration	N/A	N/A
Statistik	http://www.statistik.de/	Switzerland	3018	Bern	Morgenstrasse 129	Company	Business, Internet, Personal, Scientific	N/A	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
Strategic Feed	http://www.strategicfeed.com/	France	92 088	Paris	Place de la Pyramide 5	Company	Business, Personal	Tertiary Sector, Quaternary Sector	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
Suhoi Centre of Expertise in the Social Sciences	http://www.suhoi.com/	Switzerland	CH-1015	Lausanne	Bâtimet Gloggi	Research	Personal, Scientific	Tertiary Sector, Quaternary Sector	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
Synthetic Partners	http://www.syntheticpartners.com/en/	Spain	28006	Madrid	Calle Velázquez 92	Company	Business, Internet, Tertiary Sector	N/A	N/A	N/A	Finance, Banking, Insurance, Health and Social Care, IT Services, Public Admin	N/A	N/A
TeamGig	http://www.teamgig.com/	The Netherlands	NL-1022 AB	Amsterdam	Nieuwendammerdijk 28A-17	Company	Business, Internet, Secondary Sector, Tertiary Sector	N/A	N/A	N/A	Finance, Banking, Insurance, Retail/Trade	Research and Development (e.g. Tech)	N/A
They say	http://www.theysay.co.uk/	UK	EC2V 3ND	London	Canal 24	Company	Business, Internet	Secondary Sector	N/A	N/A	N/A	Finance, Banking, Insurance, Retail/Trade	N/A
Translight	http://www.translight.com/	Germany	D-01307	Dresden	Tatzberg 47.51	Company	Business, Internet, Tertiary Sector, Quaternary Sector	N/A	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
Transliterator	http://www.transliterator.com/	Belgium, The Netherlands, Germany, USA	N/A	N/A	N/A	Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
UK Data Archive	http://www.ukda.ac.uk/	UK	CO4 3UJ	Essex	Wivenhoe Park	Research	All	Tertiary Sector, Quaternary Sector	N/A	N/A	N/A	Finance, Banking, Insurance, Research and Development (e.g. Tech)	N/A
University of Bristol	http://www.bristol.ac.uk/	UK	BS8 1TH	Bristol	Tyndall Avenue	Research	Scientific	Quaternary Sector	N/A	N/A	N/A	Education, Research and Development (e.g. Tech)	N/A
USRO	http://www.usro.com/	Denmark	1190	Brussels	Boulevard de l'Atomité	Company	Business	N/A	N/A	N/A	N/A	N/A	N/A
Varkit	http://www.varkit.com/	USA	NI 07310-188	Jersey City	545 Washington Boulevard	Company	All	N/A	N/A	N/A	N/A	Finance, Banking, Insurance, Research and Development (e.g. Tech)	N/A
VistTrials	http://www.vist.com/	USA	UT 84113-9555	Lake City	N/A	Company	All	Quaternary Sector	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Xerox Lab	http://www.xerox.com/	USA	38240	Maylan	6 chemin de Maupertuis	Research	Business, Internet, Tertiary Sector	N/A	N/A	N/A	Finance, Banking, Insurance, Research and Development (e.g. Tech)	N/A	N/A

