Social Mining & Big Data Ecosystem

SoBigData

RESEARCH INFRASTRUCTURE

Project Acronym	***SoBigData***
Project Title	***SoBigData Research Infrastructure*** ***Social Mining & Big Data Ecosystem***
Project Number	***654024***
Deliverable Title	***Data Management report***
Deliverable No.	***D8.1***
Delivery Date	***01 December 2015***
Authors	***Valerio Grossi (CNR), Vittorio Romano (CNR), Roberto Trasarti (CNR)***

SoBigData receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654024

DOCUMENT INFORMATION

PROJECT	
Project Acronym	SoBigData
Project Title	SoBigData Research Infrastructure Social Mining & Big Data Ecosystem
Project Start	1st September 2015
Project Duration	48 months
Funding	H2020-INFRAIA-2014-2015
Grant Agreement No.	654024
DOCUMENT	
Deliverable No.	D8.1
Deliverable Title	Data Management report
Contractual Delivery Date	01 December 2015
Actual Delivery Date	01 December 2015
Author(s)	Valerio Grossi (CNR), Vittorio Romano (CNR), Roberto Trasarti (CNR)
Editor(s)	Valerio Grossi (CNR), Vittorio Romano (CNR)
Reviewer(s)	Gerhard Gossen (LUH), Beatrice Rapisarda (CNR), Thomas Risse (LUH)
Contributor(s)	Paolo Manghi (CNR), Pasquale Pagano (CNR), Leonardo Candela (CNR), All
Work Package No.	WP8
Work Package Title	JRA1_Big Data Ecosystem
Work Package Leader	CNR
Work Package Participants	CNR, USFD, UNIPI, FRH, UT, IMT, LUH, KCL, SNS, AALTO, ETHZ
Dissemination	PU
Nature	Other
Version / Revision	V1.0
Draft / Final	Final
Total No. Pages (including cover)	19
Keywords	Data Management, datasets, metadata, dataset census

DISCLAIMER

SoBigData (654024) is a Research and Innovation Action (RIA) funded by the European Commission under the Horizon 2020 research and innovation programme.

SoBigData proposes to create the Social Mining & Big Data Ecosystem: a research infrastructure (RI) providing an integrated ecosystem for ethic-sensitive scientific discoveries and advanced applications of social data mining on the various dimensions of social life, as recorded by "big data". Building on several established national infrastructures, SoBigData will open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research.

This document contains information on SoBigData core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as SoBigData Board members. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The document has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the SoBigData Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. 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 the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 27 member states of the European 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/).

GLOSSARY

ABBREVIATION	DEFINITION
RI	Research Infrastructure
WP	Work Package
[TSMM]	Text and Social Media Mining
[SNA]	Social Network Analysis
[HMA]	Human Mobility Analytics
[WA]	Web Analytics
[VA]	Visual Analytics
[SD]	Social Data
TA	Transnational Access
VA	Virtual Access
NDA	Non Disclosure Agreement

TABLE OF CONTENT

DOCUMENT INFORMATION 2
DISCLAIMER 3
GLOSSARY 4
TABLE OF CONTENT 5
DELIVERABLE SUMMARY 6
EXECUTIVE SUMMARY 7
1 Relevance to SoBigData 8
1.1 Purpose of this document 8
1.2 Relevance to project objectives 8
1.3 Relation to other workpackages 8
1.4 Structure of the document 8
2 Data Management 9
2.1 Relevant metadata fields for dataset census 9
2.2 Dataset Wiki - Input Form description 11
3 First Census of dataset of the Consortium 12
4 Conclusion 13
Appendix A. The complete list of the first census datasets available at 30 Nov 2015 14
REFERENCES 18

DELIVERABLE SUMMARY

This deliverable gives a complete description of all activities related to to provide an ongoing and up to date wiki containing the description of the available datasets in the consortium.

- **Section 1**: provides an introduction of the aim of the deliverable and its relation with the other work packages.
- **Section 2**: reports on a first study on data management plan. In particular, Section 2.1 reports a first analysis on the metadata required for describing a dataset. Section 2.2 presents the web form and the wiki page, where the description of the available datasets can be accessed or a new data set can be inserted.
- **Section 3**: provides the results of a first dataset census among the partners at Month 3.We recall that the list of the available datasets is an ongoing wiki updated through the project lifetime.

EXECUTIVE SUMMARY

This deliverable describes a web content that provides an ongoing and up to date wiki containing the description of the datasets available in the consortium. The description includes statistics, metadata, sharing policies and archiving technologies as well as the preservation provisions and lifespan. For doing that a set of relevant metadata has been defined in order to provide an homogeneous view of the datasets. The defined set of metadata will be useful also for making the datasets available into the RI. In this perspective, this deliverable represents a first definition of the metadata for describing a dataset that will be available into the RI.

Furthermore, this document presents the web form to insert a description of a new dataset, and the wiki page containing the list of the datasets available among the partners,. The proposed wiki page shows a set of relevant information, such as the name of dataset, the accessibility policy, the reference partner.

Finally, this document provides a first census of the datasets available in the consortium.

1 RELEVANCE TO SOBIGDATA

1.1 PURPOSE OF THIS DOCUMENT

This deliverable outlines a description of a web content that provides an ongoing and up to date wiki containing the description of the datasets available in the consortium. The description includes statistics, metadata, sharing policies and archiving technologies as well as the preservation provisions and lifespan. For doing that a set of relevant metadata has been defined in order to provide an homogeneous view of the datasets and to define a set of metadata that can be useful also for making the datasets available into the RI.

1.2 RELEVANCE TO PROJECT OBJECTIVES

This document shows a first census of the datasets available in the consortium. In this perspective, this deliverable represents a first definition of the metadata for describing a dataset that will be available into the RI.

1.3 RELATION TO OTHER WORKPACKAGES

This work is related to WP10. It is related to tasks T10.1 and T10.2 where all the resources that will be available in the e-infrastructure are defined. Datasets are one of the resources that will be integrated into the infrastructure. All the analysis on the relevant metadata for the data sets description and all the information inserted at this stage will be reused and enhanced during the actual integration of the datasets into the e-infrastructure.

1.4 STRUCTURE OF THE DOCUMENT

The document is categorised into 3 main sections: Section 2 provides a first overview of the data management in the SoBigData project. It is important to recall that this document represent only a first census of the all datasets available in the consortium. The list and description of the datasets will be continuously updated through time. Moreover, the actual data management plan will be developed inside WP10, when several guidelines will be delivered for integrating, and accessing data sets through the e-infrastructure. Towards this goal, Section 2.1 proposes a first analysis of the relevant fields required for the metadata to be used for describing a dataset. Section 2.2 shows the actual wiki document for access to the description of metadata. Finally, Section 3 shows the first census of the data sets available, including some statistics about datasets.

2 DATA MANAGEMENT

The purpose of the data management plan is to define policies that will guide the partners in the collection, description, preservation and sharing of their data sets for VA and TA.

Research on social mining relies on massive data sets of digital traces of human activities. Many big data sets are already available at the proposer's labs including call graphs from mobile phone call data, networks crawled from many online social networks, including Facebook and Flickr, transaction micro-data from diverse retailers, query logs both from search engines and e-commerce, society-wide mobile phone call data records, GPS tracks from personal navigation devices, survey data about customer satisfaction or market research, large Web archives, billions of tweets, and data from location-aware social networks. The partners will make such data available for collaborative research by adopting various strategies, ranging from sharing the open data sets with the scientific community at large, to sharing the data with disclosure restriction within the consortium, also on a bilateral basis, or allowing data access within secure environments at each local installation. The access under VA and TA offered in SoBigData will concern both existing and newly collected datasets. The access trough VA will be granted for all those datasets whose policies allow open diffusion; conversely, for all the data sets whose access is restricted due to licensing restrictions, access will be provided only through TA.

2.1 RELEVANT METADATA FIELDS FOR DATASET CENSUS

In accordance to the requirements of WP10 for e-infrastructure integration (tasks T10.1 "e-infrastructure interoperability" and T10.2 "Integration to the e-infrastructure"), Table 1 reports the relevant fields required for this initial dataset census. The table provides a brief description of all fields for describing the datasets inside the SoBigData consortium.

Most of the fields are coded using the DataCite standard [1]. We used Dublin Core Metadata Initiative [2], when it was not possible to employ the DataCite standard. Only few fields are proprietary and explicitly defined for the SoBigData project, for example the six thematic clusters recognized inside the project. Finally, we have highlighted all the fields that are mandatory, i.e. the name of the datasets, or the accessibility.

The analysis of the required fields takes into account the integration of the datasets into the e-infrastructure. All the fields in Table 1 will be reused for defining the metadata structure required also for datasets integration into the infrastructure.

id	*field name*	*description*	*standard*	*mandatory*
1	**Name**	name of the dataset	**DataCite:Title**	***yes***
2	**Description**	All additional information that does not fit in any of the other categories. May be used for technical information.	**DataCite:Description (descriptionType=abstract)**	***yes***
3	**Identifier**	The Identifier is a unique string that identifies a resource. The value should be a DOI or a similar globally unique identifier.	**DataCite:Identifier**	
4	**Creator**	The main researchers involved in producing the data, or the authors of the	**DataCite:Creator**	***yes***

		publication, in priority order.		
5	**Creation Date**	The date of creation of the dataset.	**DataCite:Date (dateType=Created)**	***yes***
6	**Providing Laboratory**	The SoBigData Partner Laboratory who provides the dataset e.g. "*KDD Lab ISTI CNR*" or "*SNS*"	**DataCite:Contributor (contributorType=distributor)**	***yes***
7	**Manifestation Type**	Use “Original” for collections of data produced and kept in local infrastructure by the data provider, “Replica” for a copy of data in remote sites (e.g. DBPL) or “Virtual” if the collection is accessible in streaming from remote sites	**Custom**	***yes***
8	**Publisher**	The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role.	**DataCite:Publisher**	***yes***
9	**Publication Date**	The year when the data was or will be made publicly available. This field is not mandatory, since proprietary dataset do not have a publication date.	**DataCite:PublicationYear**	
10	**Contact Person**	The contact person for the dataset.	**DataCite:Contributor (contributorType=ContactPerson)**	***yes***
11	**Thematic Cluster**	The six thematic clusters described in the project proposal	**DataCite:Subject**	***yes***
12	**Semantic Coverage**	Comma separated list of tags about the dataset. Allowed values, examples, other constraints: Tagging e.g. people, cities, transports...	**DataCite:Subject**	***yes***
13	**Time Coverage**	The start and end dates of the period to which the data is relevant.	**DublinCore:coverage.temporal**	
14	**Geo Location**	Spatial region (WGS 84, World Geodetic System) or named place where the data was gathered or about which the data is focused.	**DataCite:GeoLocation**	
15	**Processing Degree**	Use "Primary" for raw data or "Secondary" for processed data	**Custom**	***yes***
16	**Rights**	Any rights information for this resource	**DataCite:Rights**	***yes***
17	**Accessibility**	The accessibility criteria of the dataset in SoBigData	**Custom**	***yes***
18	**Privacy**	Does the dataset contain sensitive information? What are the policies for preserving the privacy of the users?	**Custom**	
19	**Disk Size**	The size of the dataset on disk, in MB.	**DataCite:Size**	

20	**Format**	Technical format of the resource.	**DataCite:Format**	***yes***
21	**Language**	The primary language of the resource.	**DataCite:Language**	

Table 1. Relevant fields and their standard for Dataset Wiki

2.2 DATASET WIKI - INPUT FORM DESCRIPTION

In order to guarantee a fast and smart way for describing and conducting a census among the partners, and to provide a wiki-type document where all the datasets description can be easy visualized; a web dataset catalog, and a web form has been developed into SoBigData.eu web site[1]. Figure 1 shows a snapshot of the web page where the data catalog can be viewed.

For adding a new dataset, you have to click on "***Add new dataset***". For adding a new dataset a simple form is available, and after all the information required are inserted the description of new dataset is added into the site, and than can be visualized. The fields required for the input form are the one described in Table 1.

Considering Figure 1, by clicking on "***Dataset List***" the site presents, in an interactive way, the list of the datasets available among the partners, showing a set of relevant information, such as the name of dataset, the accessibility polity. Currently, this page enable the user to filter, order and access to the description of a data set. At this stage, we recall that the this documents represents only an introduction on the information and the presentation of the wiki containing the description of the available datasets in the consortium, this wiki is ongoing and continuously up to date through time.

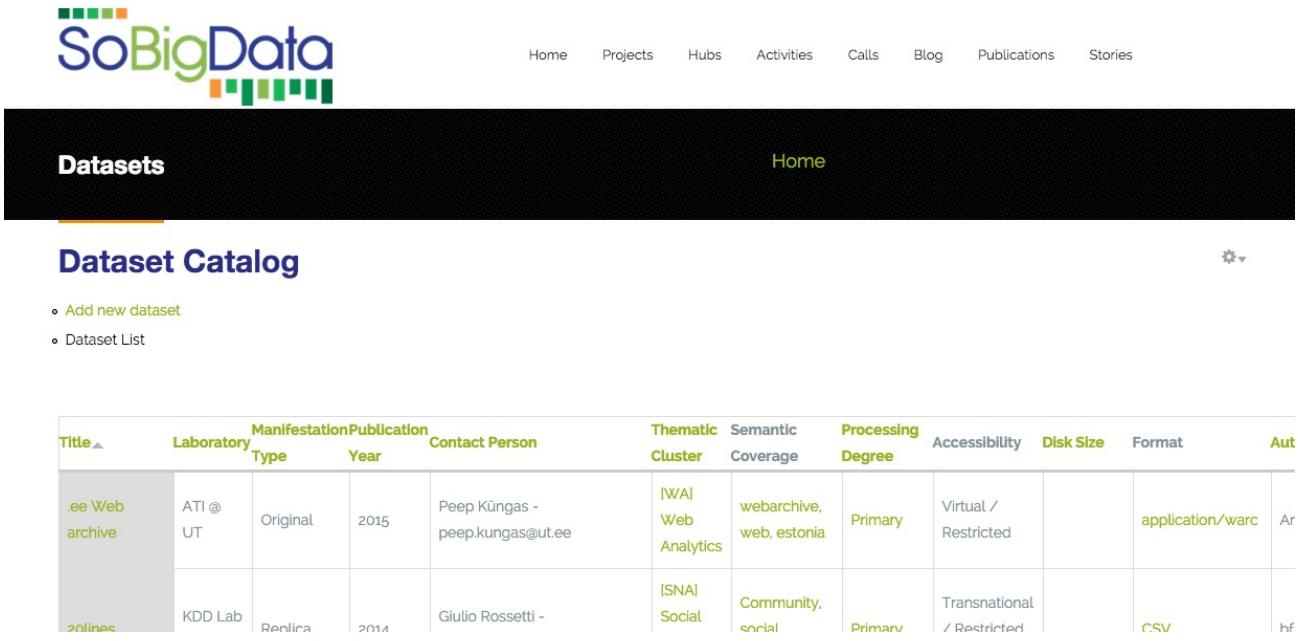

Figure 1. The data set catalog

[1] http://www.sobigdata.eu/private/ouhnr3inebchcerhncurhebcnhn8749892/datasets.

3 FIRST CENSUS OF DATASET OF THE CONSORTIUM

This final section proposes a first census of the datasets available in the consortium. This represents only an initial analysis of the resources available into the project. The list of datasets is ongoing and up-to-date through time. At the end of this first census, 30 Nov 2015 the consortium shows:

- 63 datasets from 10 partners, five thematic clusters covered with the following distribution[2]:
    - [HMA] Human Mobility Analytics: 15 datasets
    - [SD] Social Data: 6 datasets
    - [SNA] Social Network Analysis: 15 datasets
    - [TSMM] Text and Social Media Mining: 21 datasets
    - [VA] Visual Analytics: 0 datasets
    - [WA] Web Analytics: 6 datasets

- accessibility: a dataset can be accessed with virtual and/or transnational way. This information is integrated considering if a dataset is: public for public data e.g. open data; restricted for data available under specific restrictions e.g. NDA. Finally, private is meant for dataset that cannot be accessed by the user directly but only through APIs, services or views that provide only aggregated information or analysis. Private datasets can also be used by the partners in order to perform custom analysis for the user. It is worth to notice that multiple choices can be selected for a dataset. For example a dataset can be marked as “Virtual / Public” and “Transnational / Public”, if it is freely accessible through both Virtual and Transnational Access. Accessibility has 5 possible values with items with following distribution:
    - Virtual / Public: 10 datasets
    - Virtual / Restricted: 12 datasets
    - Transnational / Public: 8 datasets
    - Transnational / Restricted: 30 datasets
    - Private: 27 datasets

Finally, Appendix A reports the list of all datasets available into the wiki at 30 Nov 2015. For each data set, we report some relevant information such as the *accessibility*, or the *manifestation type* that shows if a dataset is a replica, i.e. has been pre-processed or transformed in some way, or is original, i.e. taken as it has been generated.

[2] All clusters (except VA) are related to specific types of data. VA instead is a cluster that indirectly applies to all datasets, since different application can use VA methods on all kind of datasets.

4 CONCLUSION

This deliverable describes an initial study about the metadata definition for describing a dataset. It aims to present on the one hand the metadata defined, and on the other hand the wiki-site for adding a new dataset and for presenting the available ones. The datasets available are continuously updated throughout the project lifetime. It important to notice that the metadata defined in this context and the description of the available datasets will be used inside WP10 for the integration of the datasets into the RI.

APPENDIX A. THE COMPLETE LIST OF THE FIRST CENSUS DATASETS AVAILABLE AT 30 NOV 2015

ID	Titlesort descending	Laboratory	Manifestation Type	Thematic Cluster	Semantic Coverage	Processing Degree	Accessibility
1	.ee Web archive	ATI @ UT	Original	[WA]	webarchive, web, estonia	Primary	Virtual / Restricted
2	20lines	KDD Lab ISTI CNR	Replica	[SNA]	Community, social network	Primary	Transnational / Restricted, Private
3	Aalto-Foursquare	Aalto Data Mining Group	Virtual	[HMA]	check-ins, foursquare, location based social networks	Secondary	Virtual / Restricted
4	Aalto-Twitter	Aalto Data Mining Group	Virtual	[TSMM]	twitter, tweets	Secondary	Virtual / Restricted
5	Archive Twitter Dataset	ETH Zurich	Replica	[TSMM]	tweets text and metadata	Primary	Transnational / Public
6	Articles and comments of major Estonian newspapers	ATI @ UT	Replica	[TSMM]	news articles, comments, persons, organizations	Secondary	Virtual / Restricted
7	BrightKite	KDD Lab ISTI CNR	Replica	[SNA]	mobility, checkin, people	Primary	Virtual / Public
8	ClueWeb 2009	HPC Lab ISTI CNR	Replica	[WA]	web pages, information retrieval, search	Primary	Transnational / Restricted, Private
9	ClueWeb 2012	HPC Lab ISTI CNR	Replica	[WA]	web pages, information retrieval, search	Primary	Transnational / Restricted, Private
10	Collection of texts for data compression	A^3 lab UNIPI	Original	[TSMM]	texts	Primary	Transnational / Public
11	Coop	KDD Lab ISTI CNR	Replica	[SD]	purchasing data, people behavior	Primary	Transnational / Restricted, Private
12	CoPhIR	HPC Lab ISTI CNR	Original	[TSMM]	photos, image features	Secondary	Virtual / Restricted
13	Corpora of web crawls in English, French,	USFD	Original	[TSMM]	web	Primary	Virtual / Restricted,

	Italian and German						Transnational / Restricted
14	DBLP	KDD Lab ISTI CNR	Replica	[SNA]	people, bibliography	Primary	Virtual / Public
15	DE webarchive	LUH	Original	[WA]	webarchive, web, germany	Primary	Virtual / Restricted, Transnational / Restricted
16	e-MID	SNS	Replica	[SNA]	Finance, Interbank market	Secondary	Private
17	EMID Data	IMT	Original	[SD]	Interbank market	Primary	Transnational / Restricted
18	English documents annotated for IE tasks	USFD	Original	[TSMM]	web, conversation, newswire	Secondary	Transnational / Restricted
19	English Newswire	USFD	Original	[TSMM]	newswire	Primary	Transnational / Restricted
20	Estonian public sector Web services	ATI @ UT	Replica	[SNA]	Web services, service providers, service consumers	Secondary	Virtual / Public
21	Facebook - EuroSys '09	UI IIT CNR	Original	[SNA]	facebook, social graph, interaction graph	Primary	Virtual / Public
22	Facebook - WOSN '09	UI IIT CNR	Original	[SNA]	facebook, social graph, interaction graph	Primary	Virtual / Public
23	Flickr and Wikipedia Turism Trajectories	HPC Lab ISTI CNR	Original	[HMA]	trajectories, points of interest, mobility	Secondary	Transnational / Public
24	Formal network of Estonian companies and board members	ATI @ UT	Replica	[SNA]	people, organizations, board membership	Secondary	Private
25	German Academic Web	LUH	Original	[WA]	germany, academia, webarchive	Primary	Virtual / Restricted, Transnational / Restricted
26	Google+	KDD Lab ISTI CNR	Replica	[SNA]	social network	Primary	Virtual / Public
27	IMDb	KDD Lab ISTI CNR	Replica	[SNA]	people, movies	Primary	Virtual / Public

28	Infocamere	KDD Lab ISTI CNR	Replica	[SD]	company	Primary	Transnational / Restricted, Private
29	ISTAT Census Data Tuscany	KDD Lab ISTI CNR	Replica	[SD]	Census data	Primary	Transnational / Restricted, Private
30	Italian Twitter Dataset	A^3 lab UNIPI	Replica	[SNA]	social network	Primary	Private
31	LastFM	KDD Lab ISTI CNR	Replica	[SNA]	people, music, webradio	Primary	Transnational / Public
32	Linguistically annotated corpora of IRC chats, reviews, Q&A, email, blogs and comments, newsgroups	USFD	Original	[TSMM]	chats, reviews, blogs, comments, newsgroups	Secondary	Transnational / Restricted
33	Mobile Miner App Data	Dept of Digital Humanities Kings College London	Original	[SD]	smartphones, mobile apps	Primary	Virtual / Restricted
34	OctoCalabria 2012 10 e 07	KDD Lab ISTI CNR	Replica	[HMA]	mobility, GPS	Primary	Transnational / Restricted, Private
35	OctoMestre 2010 08	KDD Lab ISTI CNR	Replica	[HMA]	mobility, GPS	Primary	Transnational / Restricted, Private
36	OctoMilano 2007 04	KDD Lab ISTI CNR	Replica	[HMA]	mobility, GPS	Primary	Transnational / Restricted, Private
37	OctoPisa	KDD Lab ISTI CNR	Replica	[HMA]	mobility, GPS, people	Primary	Transnational / Restricted, Private
38	Octoscana 2011 05	KDD Lab ISTI CNR	Replica	[HMA]	mobility, GPS	Primary	Transnational / Restricted, Private
39	Orange D4D	KDD Lab ISTI CNR	Replica	[HMA]	mobile phone data, cdr, mobility	Secondary	Transnational / Restricted, Private
40	Pisa Airport statistics	KDD Lab ISTI CNR	Replica	[HMA]	tranport, airplain, mobility	Secondary	Transnational / Public
41	Pisa Hotel statistics	KDD Lab ISTI CNR	Replica	[HMA]	people, tourism, hotel	Primary	Transnational / Restricted, Private

42	Pizza&Chili	A^3 lab UNIPI	Replica	[TSMM]	texts	Primary	Transnational / Public
43	Product Reviews Rating Datasets	NeMIS Lab ISTI CNR	Replica	[TSMM]	product reviews	Primary	Private
44	Query Log MSN RFP 2006	HPC Lab ISTI CNR	Replica	[WA]	query log, clicks, search, web mining	Primary	Transnational / Restricted, Private
45	Query-Based Twitter Dataset	ETH Zurich	Original	[TSMM]	tweets text and metadata	Primary	Transnational / Public
46	Russell 3000	SNS	Replica	[SNA]	Finance, stock exchange	Secondary	Virtual / Restricted
47	SentiWordNet	NeMIS Lab ISTI CNR	Original	[TSMM]	semantic enrichment	Secondary	Virtual / Public
48	SMAPH Annotated query dataset	A^3 lab UNIPI	Replica	[TSMM]	annotated search engine queries	Secondary	Private
49	Social Banks Tracker Twitter Dataset	A^3 lab UNIPI	Virtual	[TSMM]	tweets	Primary	Private
50	Strava	KDD Lab ISTI CNR	Replica	[SD]	mobility, people behavior	Primary	Transnational / Restricted, Private
51	TAGME Datasets	A^3 lab UNIPI	Replica	[TSMM]	annotated tweets and wikipedia pages	Secondary	Transnational / Public
52	TagMyDay (dayTag)	KDD Lab ISTI CNR	Original	[HMA]	mobility, semantic enrichment	Primary	Transnational / Restricted, Private
53	Thomson Reuters	SNS	Replica	[SNA]	Finance, stock exchange	Primary	Virtual / Restricted
54	Trenitalia	KDD Lab ISTI CNR	Replica	[HMA]	mobility, public transport	Primary	Transnational / Restricted, Private
55	TripAdvisor Annotated Dataset	NeMIS Lab ISTI CNR	Original	[TSMM]	semantic enrichment	Secondary	Virtual / Restricted
56	Twitter Gardenhose @ USFD	USFD	Replica	[TSMM]	raw data	Primary	Transnational / Restricted
57	Twitter Stream - Gardenhose Daily Access	HPC Lab ISTI CNR	Original	[TSMM]	social media, microblogging, tweets	Primary	Transnational / Restricted, Private
58	US Banks balance	SNS	Replica	[SNA]	Banks, Finance,	Secondary	Virtual / Public

	sheets				portfolios		
59	Web 1T 5-gram, English + 10 EU languages	USFD	Original	[TSMM]	web	Primary	Transnational / Restricted
60	Wikipedia Graph	HPC Lab ISTI CNR	Virtual	[TSMM]	Wikipedia graph	Primary	Virtual / Public
61	Wind 2012 02	KDD Lab ISTI CNR	Replica	[HMA]	mobile phone data, cdr, mobility	Primary	Transnational / Restricted, Private
62	Wind 2013 10	KDD Lab ISTI CNR	Replica	[HMA]	mobile phone data, cdr, mobility	Primary	Transnational / Restricted, Private
63	Wind 2014 03	KDD Lab ISTI CNR	Replica	[HMA]	mobile phone data, cdr, people, mobility	Primary	Transnational / Restricted, Private

REFERENCES

[1] DataCite - International Data Citation, “DataCite Metadata Schema for the Publication and Citation of Research Data “, Version 3.1, August 2015 doi:10.5438/0010, https://www.datacite.org/sites/default/files/document/DataCite-MetadataSchema_V31_Final_8-24-2015_0.pdf, 2015.

[2] Dublin Core Metadata Initiative Specifications http://dublincore.org/specifications/, 2015.