“Building the Next Generation Personal Data Platforms”
G.A. n. 871370

DELIVERABLE D7.1

Data Management Plan

H2020-EU-2.1.1: PIMCity

Project No. 871370

Start date of project: 01-12-2019

Duration: 30 months

Revision: 01

Date: 24-05-2020

Document Information

Document Name: Data Management Plan

WP7 – Title: Data management and legal and ethical requirements

Task 7.1

Revision: 01

Revision Date: 09/06/2020

Author: KUL and all Partners

Dissemination Level

Project co-funded by the EC within the H2020 Programme		
PU	Public	☑
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

(Tick the corresponding dissemination level of the deliverable according to Annex I).

Approvals

	Name	Entity	Date	Visa
Author	**Alessandro Bruni, Aleksandra Kuczerawy, Viltė Kristina Steponėnaitė**	**KUL**	**31-05-2020**	☑
WP Leader	**Viltė Kristina Steponėnaitė**	**KUL**	**31-05-2020**	☑
Coordinator	**Marco Mellia**	**POLITO**	**03-06-2020**	☑

Document history

Revision	Date	Modification
Version 1	31-05-2020	V1. Review of the data management plan (excluding Annex F) and Annexes A-E
Version 2	31-05-2020	Review of the data management plan (including Annex F)

LIST OF ABBREVIATIONS

Abbreviation	Meaning
CDSM	Directive on Copyright and related rights in the Digital Single Market
CJEU	Court of Justice of the European Union
Database Directive	Directive on the legal protection of databases
Deliverable D7.2	Deliverable D7.2 of the PIMCity project
DMP	Data Management Plan
DPIA	Data protection impact assessment
DPO	Data protection officer
EU	European Union
FAIR data	Findable, accessible, interoperable and re-usable data
GDPR	General Data Protection Regulation
ePD	e-Privacy Directive
InfoSoc	Directive on the harmonisation of certain aspects of copyright and related rights in the information society
KUL	KU Leuven – CiTiP
NPD Regulation	Regulation on the free flow of non-personal data
PIMCity project	Horizon2020 project PIMCity (Building the next generation Personal Data Platforms), Grant Agreement No. 871370
TFEU	Treaty on the Functioning of the European Union
Trade Secrets Directive	Directive on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure

TABLE OF CONTENTS

INTRODUCTION 8
PIMCITY DATA MANAGEMENT PLAN 9
1. DATA SUMMARY 12
2. FAIR DATA 17
2.1. Making data findable, including provisions for metadata 17
2.2. Making data openly accessible 17
2.3. Making data interoperable 19
2.4. Increase data re-use (through clarifying licences) 19
3. ALLOCATION OF RESOURCES 20
4. DATA SECURITY 20
5. ETHICAL ASPECTS 21
6. OTHER 21
ANNEX A – DATA PROTECTION IMPACT ASSESSMENT 21
ANNEX B – GUIDELINES FOR CONSENT MANAGEMENT 25
ANNEX C – GUIDELINES FOR PRIVACY POLICIES 29
ANNEX D – TEMPLATE INFORMED CONSENT FORM 31
ANNEX E – TEMPLATE PRIVACY POLICY 32
ANNEX F – PARTNERS' INDIVIDUAL INPUT FOR THE DATA MANAGEMENT PLAN 37
1. DATA SUMMARY 37
1.1. Politecnico di Torino 37
1.2. NEC Laboratories Europe 38
1.3. Ermes Cyber Security 38
1.4. IMDEA Networks 39
1.5. Universidad Carlos III de Madrid 40
1.6. Telefónica Investigación y Desarrollo 41
1.7. Fastweb 41
1.8. LSTech ESPANA 42
1.9. KU Leuven – CiTiP 43
1.10. Asociación de Usuarios de Internet 44
1.11. Big Data Analytics 45
1.12. CLIQZ 46
2. FAIR DATA 46
a. Making data findable, including provisions for metadata 46

2.1.1. Politecnico di Torino 46
2.1.2. NEC Laboratories Europe 47
2.1.3. Ermes Cyber Security 47
2.1.4. IMDEA Networks 48
2.1.5. Universidad Carlos III de Madrid 48
2.1.6. Telefónica Investigación y Desarrollo 49
2.1.7. Fastweb 49
2.1.8. LSTech ESPANA 50
2.1.9. KU Leuven – CiTiP 51
2.1.10. Asociación de Usuarios de Internet 51
2.1.11. Big Data Analytics 52
1.13. CLIQZ 52
b. Making data openly accessible 53
2.2.1. Politecnico di Torino 53
2.2.2. NEC Laboratories Europe 54
2.2.3. Ermes Cyber Security 55
2.2.4. IMDEA Networks 57
2.2.5. Universidad Carlos III de Madrid 58
2.2.6. Telefónica Investigación y Desarrollo 59
2.2.7. Fastweb 60
2.2.8. LSTech ESPANA 62
2.2.9. KU Leuven – CiTiP 63
2.2.10. Asociación de Usuarios de Internet 64
2.2.11. Big Data Analytics 65
1.14. CLIQZ 66
c. Making data interoperable 68
2.3.1. Politecnico di Torino 68
2.3.2. NEC Laboratories Europe 68
2.3.3. Ermes Cyber Security 69
2.3.4. IMDEA Networks 69
2.3.5. Universidad Carlos III de Madrid 70
2.3.6. Telefónica Investigación y Desarrollo 70
2.3.7. Fastweb 71
2.3.8. LSTech ESPANA 71
2.3.9. KU Leuven – CiTiP 72

2.3.10. Asociación de Usuarios de Internet 72
2.3.11. Big Data Analytics 73
2.3.12. CLIQZ 73
d. Increase data re-use (through clarifying licences) 74
2.4.1. Politecnico di Torino 74
2.4.2. NEC Laboratories Europe 74
2.4.3. Ermes Cyber Security 75
2.4.4. IMDEA Networks 75
2.4.5. Universidad Carlos III de Madrid 76
2.4.6. Telefónica Investigación y Desarrollo 77
2.4.7. Fastweb 77
2.4.8. LSTech ESPANA 77
2.4.9. KU Leuven – CiTiP 78
2.4.10. Asociación de Usuarios de Internet 78
2.4.11. Big Data Analytics 79
2.4.12. CLIQZ 80
3. ALLOCATION OF RESOURCES 80
4. DATA SECURITY 81
4.1. Politecnico di Torino 81
4.2. NEC Laboratories Europe 81
4.3. Ermes Cyber Security 81
4.4. IMDEA Networks 81
4.5. Universidad Carlos III de Madrid 82
4.6. Telefónica Investigación y Desarrollo 82
4.7. Fastweb 82
4.8. LSTech ESPANA 83
4.9. KU Leuven – CiTiP 83
a. Asociación de Usuarios de Internet 83
4.10. Big Data Analytics 83
4.11. CLIQZ 84
5. ETHICAL ASPECTS 84
5.1. Politecnico di Torino 84
5.2. NEC Laboratories Europe 84
5.3. Ermes Cyber Security 84
5.4. IMDEA Networks 85

5.5. Universidad Carlos III de Madrid 85
5.6. Telefónica Investigación y Desarrollo 85
5.7. Fastweb 85
5.8. LSTech ESPANA 85
5.9. KU Leuven – CiTiP 85
5.10. Asociación de Usuarios de Internet 85
5.11. Big Data Analytics 86
5.12. CLIQZ 86
6. OTHER 86
6.1. Politecnico di Torino 86
6.2. NEC Laboratories Europe 86
6.3. Ermes Cyber Security 86
6.4. IMDEA Networks 86
6.5. Universidad Carlos III de Madrid 87
6.6. Telefónica Investigación y Desarrollo 87
6.7. Fastweb 87
6.8. LSTech ESPANA 87
6.9. KU Leuven – CiTiP 87
6.10. Asociación de Usuarios de Internet 87
6.11. Big Data Analytics 87
6.12. CLIQZ 88

INTRODUCTION

The PIMCity Data Management Plan (hereinafter as the DMP) sets out data management plan for the PIMCity project and reflects the consortium's comprehensive approach and joint efforts towards making research data findable, accessible, interoperable and re-usable (further as FAIR). The DMP describes the data management life cycle for the data to be collected, processed and/or generated by the PIMCity project. As part of making research data FAIR, the DMP provides the information on the handling of research data during and after the end of the project. The DMP indicates what data will be collected, processed and/or generated, which methodology & standards will be applied, whether data will be shared/made open access and how data will be curated & preserved (including after the end of the project)[1]. Content of the DMP is limited to, i.e. based solely on the information provided by the project partners Politecnico di Torino, NEC Laboratories Europe, Ermes Cyber Security, IMDEA Networks, Universidad Carlos III de Madrid, Telefónica Investigación y Desarrollo, Fastweb, LSTech ESPANA, KU Leuven – CiTiP, Asociación de Usuarios de Internet, Big Data Analytics, CLIQZ as indicated in the relevant sections. The DMP is drafted in accordance with the Guidelines on FAIR Data Management in Horizon 2020 as of 26th July, 2016, issued by the European Commission Directorate-General for Research & Innovation.

The document also includes the data protection impact assessment, guidelines for consent management, privacy policies, data processing agreements and templates of informed consent forms, data processing agreements and controller-to-controller agreements that shall be adjusted by the partners' on a case by case basis taking into account particular details.

STRUCTURE

The DMP consists of two parts. The first part clarifies the general approach towards data management as adopted by the project partners. The second part consists of partners' individual inputs which clarify the approach initially suggested by each partner (Annex F). The data protection impact assessment, guidelines and templates are added as Annexes A-E.

DISCLAIMERS

The current version of the DPM shall not suggest that it covers all of the data management details exhaustively. Some details shall be clarified and agreed upon by the PIMCity project partners at the later stages. Currently, inputs from some partners are missing, e.g. IAB Spain as a new partner of the PIMCity project has not provided its input yet. The DMP is a living document and shall be updated during the project taking into account to the possibly significant changes, including but not limited the use of new data, changes in consortium policies, changes in consortium composition and external factors (e.g. new consortium members joining or existing members leaving).[2]

Taking into account the information provided by the PIMCity partners, the legal implications related to the content of the DMP shall be noted. The DMP may contain the PIMCity

[1] European Commission Directorate-General for Research & Innovation. Guidelines on FAIR Data Management in Horizon 2020 as of 26th July, 2016, p. 4.
[2] Ibid, p. 5.

partners' copyrighted material that may not be used without permission. The information provided in this document may be confidential and may not be disclosed except under a consortium agreement. The commercial use the information provided in this document may require a license.

The information provided in the DMP, guidelines and templates (Annexes A-E) does not constitute legal advice. Any user of this information uses it at its sole risk and liability.

PIMCITY DATA MANAGEMENT PLAN

In general, the PIMCity project is expected to collect and/or generate at least four broad categories of data as provided in the table below. The information provided further in the DMP shall supplement and refine this information. The DMP shall elaborate on the management of these broad categories of data and shall reveal in detail the relation of each dataset with the PIMCity project's exploitable outputs to gain a clear understanding of the limitations of the OA as well as the impact on the exploitation and dissemination strategies.

Category	Format	Means for OA	Curation and cost allocation
Documents and dissemination materials: Includes deliverables, reports, demonstrations, dissemination and communication	Common text, image or video formats (.pdf, .docx, .jpeg, .mov, .avi, etc.)	Self-archive on website; green scientific publications + OPENAIRE repositories	Technical coordinator (T8.2); dissemination manager (T6.1); peer review: scientific journal panels
Computer software: including software applications (in binary form), libraries in the form of SDKs, plugins and respective source code	Binary format, ZIP archives; Source code in common files such as C, CPP, etc.	GitHub	Technical coordinator (T8.2); innovation manager: exploitability and license schemes (T6.4)
Research data and metadata: materials and datasets resulting from the implementation of the developments; metadata and configuration files; bug logs and feedback logs; developer internal documentation; evaluation and opinions	Log files; text files using (.pdf, .docx, .xls, etc.)	green scientific publications + OPENAIRE repositories	Innovation manager: exploitability (T6.4); data manager: anonymization of evaluation questionnaires and opinions; conditions of pre-existent data (T7.1); dissemination manager (T6.1)
Data for evaluation: consists in materials or	Log files; files using	Green scientific	Data manager: conditions of pre-

datasets generated or collected by the project used for evaluation purposes	(.docx, .xls, etc.)	publications + OPENAIRE repositories	existent data (T7.1); dissemination manager (T6.1)

Approach to personal data

Since the PIMCity project will be dealing with privacy-preserving technologies, the project partners need to have access to personal data. Accordingly, personal data, including special categories of personal data, will be processed. In essence, personal data will be collected from participants and beta-testers, who participate on a voluntary basis and will be informed accordingly.

All of the data will be processed in accordance with the relevant EU legal requirements. To enable data subjects to effectively exercise their rights, data controllers will fulfil all of the requirements stemming from the relevant legal frameworks as defined in the deliverables of WP7, including the implementation of internal organisational and technical measures.

All personal sensitive data collected will be deleted after the end of the project, once the final demonstrations and review meeting is completed (estimated on July 2022). In any case, data will not be retained beyond October 2022. However, since the consortium commits to keep the EasyPIMS up and running for at least one year after the project final data, the project partners will specifically request end-users their willingness to keep participating and using the platform.

Organisational measures

The PIMCity partners have appointed the PIMCity data protection officer team to act as a single point of contact for data subjects wishing to exercise their rights, following the provisions of the Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data[3] (hereinafter as the GDPR) and Article 29 Working Party Guidelines on Data Protection Officers.[4] The PIMCity data protection officer team will develop a roadmap with actions to be taken if a data subject sends a request, to give a response within the timeframes provided in the relevant legal frameworks.

Technical measures

Partners will implement technical means to accommodate data subject's requests, including but not limited:

[3] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, 4.5.2016, p. 1–88.

[4] Article 29 Working Party Guidelines on Data Protection Officers, adopted on 13 December 2016, last revised and adopted on 5 April 2017 https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048, and endorsed by the European Data Protection Board which replaced the Article 29 Working Party on 25 May 2018.

- *The request to withdrawing consent*: processing of beta-testers' data shall be based on the consent of those individuals. However, beta-testers have a right to withdraw their consent at any time. If the single point of contact receives a withdrawal request, appropriate technical means shall be present to identify the data belonging to that particular individual, erase it and ensure that the data are no longer processed.

- *Right of access*: there shall be means enabling partners to extract all personal data relating to an individual and provide a copy of such data.

- *Right to erasure*: partners shall have the means to ensure the erasure of personal data.

- *Data portability*: transmitting shall be ensured using structured, commonly used and machine-readable formats, using secure methods. Controllers shall explore and assess two different and complementary ways to make portable data available: direct transmission of the overall dataset of portable data (or several extracts of parts of the global dataset) and an automated tool that allows extraction of relevant data.

More details on various mitigation measures are provided in D9.2 (confidential), coordinated by Politecnico di Torino.

Third countries

PIMCity confirms that in principle, only partners who are part of the EU/EEA Member States will be processing data. However, some internal testing of algorithms and tools will be carried out in Argentina. In such cases, it may be necessary to transfer some limited amount of data. The PIMCity partners will ensure compliance with the legal requirements relevant for such actions.

Re-use and open access

Overall, the PIMCity project partners are expected to generate a significant amount of research data with potential for re-use and verification. As identified in the PIMCity project proposal, in compliance with Responsible Research and Innovation on Open Access (hereinafter as OA), default policy is to make its data publicly available through public copyright licenses (e.g. Creative Commons, etc.). Possible solution can be, among others, archiving it on the project website and OPENAIRE compliant repositories. However, with regard to certain data PIMCity partners may need to apply OA restrictions. The latter may stem from, including but not limited, (i) confidentiality and intellectual property protection of certain deliverables, datasets and outputs (e.g. underlying algorithmic and methods susceptible of being patented); (ii) protection of personal data of persons involved in feedback collections; (iii) protection of imported rights of pre-existent, non-public datasets.

1. DATA SUMMARY	Details on (i) the purpose of the data collection/generation and its relation to the objectives of the project; (ii) the types and formats of data will the project generate/collect; (iii) the plans to re-use any existing data and, if yes, details on how; (iv) the origin of the data; the expected size of the data; (v) to whom might it be useful ('data utility').
	(i) In general, data will be collected and generated for the purposes of the project which are, including but not limited, to create privacy preserving tools. In particular, the project partners aim to implement a Personal Information Management Systems (PIMS) development kit (PDK) to commoditize the complexity of creating PIMS. Second, project partners aim to design and deploy novel mechanisms to increase users' awareness: (i) the Transparency Tags (TT) which would show users essential information about the services they access, in a simple and easy to understand manner; (ii) the Personal Data Avatar (PDA) which would be intuitive means for users to control the information shared to third parties. Third, the project partners aim to demonstrate the effectiveness of the above tools by engineering EasyPIMS, the fully-fledged PIMS. Overall, the project partners partners aim to build the largest-ever transparent data marketplace implementing and demonstrating EasyPIMS with a number of end-users, collaborating with advertisers and operators in the web market. (ii) Overall, the partners intend to generate/collect these data: - information instrumental to check if and how the website or web service collects and exchanges eventual personal information; privacy tags will summarize the output of the algorithms (Politecnico di Torino); - data composed by sequences of host and stored in a mysql database (NEC Laboratories Europe); - data generated by its fleet of automatic web scrapers, which will be then processed by automatic algorithms to identify which personal data web services collect and how; the result of this analysis, the D-PM, will be provided in JSON format, and stored and distributed using state-of-the-art database technologies (Ermes Cyber Security); - existing public data available on the Internet that may eventually resemble the kind of information that the PDK or the EasyPIMS would be managing and trading; this is in general structured data, for instance anonymized

	mobility data or CDRs (call detail records) showing the mobility of people within an area, sich as a city; more details could be provided once the specific use cases to implement are clear; as of now, we will be working with this mobility information which in all cases is totally anonymized (IMDEA Networks; LSTech ESPANA); - data regarding the value/price of audiences (i.e., user profiles) in online advertising platforms (Universidad Carlos III de Madrid); - in most cases textual type with formats such as JavaScript Object Notation (JSON) and Comma Separated Values (CSV); depends on the format that the 3rd party systems provide their data (Telefónica Investigación y Desarrollo); - aggregate network traffic data (e.g. traffic share of a particular website or service, bandwidth usage over time, number of users over time); the aggregation occurs in the dimension of users, i.e.: we will not collect or share any individual user identifier, but only user counts or session counts or bandwidth per website/service, possibly over time (Fastweb); - identifying data (name, email, phone) collected exclusively to contact participants who wish to give them voluntarily; as for the formats these are collected in records and tables of mysql type databases in which the identification data are always stored encrypted (Asociación de Usuarios de Internet); - sociodemographic data (age, gender, country, marital status, professional status, level of education, country, state, city, language) for the elaboration of studies and evaluation of the use of the different tools; as for the formats these are collected in records and tables of mysql type databases in which the identification data are always stored encrypted (Asociación de Usuarios de Internet); - use and activity data (log files); as for the formats these are collected in records and tables of mysql type databases in which the identification data are always stored encrypted (Asociación de Usuarios de Internet); - information from Data Sellers and Data Buyers as JSON, responding to the transaction history in JSON format; the raw input data format for the TG varies depending on the data source, e.g. the historical geolocation data can be a KML or a JSON file; the schema generated for the data source is served as JSON (Big Data Analytics); - on the scope of the consent manager (P-CM), the data will consist on tuples <identifier_of_user_data, consent_level> for each user; the consent_level is a structure common across all users; the

	identider_of_user_data is a pointer to arbitraty user's data stored on the P-DS; on the scope of the Trading Engine (TE), the data will consist on contract, queries expresed by buyers; in the case of fullfillment of the contract, it will be stored for auditing purposed; contracts contain no personal data but pointers to all relevant actors of it, namely: identider_of_user_data, identifier_of_seller, intefider_of_buyer and other metadata relevant for the contract such as date, experiration, price, etc. (CLIQZ); - KU Leuven – CiTiP will generate research data in the form of deliverables which will be saved primarily in .docx and .pdf formats. (iii) the project partners intend to re-use existing data, however most of these data shall not be personal data; in detail: - Politecnico di Torino has historical web crawling archives that have been performed in the past; these contains a snapshot of web pages done through the time using automatic web crawlers; - NEC Laboratories Europe will probably re-use anonimyzed dataset containing traffic logs; - Ermes Cyber Security will re-use data collected in the past using its fleet of automatic web scrapers. This contains information about web sites including the code used to generate the page (e.g., HTML and Javascript) as well as logs describing the APIs executed by the browser to generate the page; - IMDEA Networks will conduct the tests by reusing public data, by downloading copies of the target datasets to local development environments and using them in the testbeds; - Universidad Carlos III de Madrid has some data collected before the start of the project, in the context of TYPES and SMOOTH EU projects and they might reuse it. It is of the same nature and type than the one they plan to collect in PIMCITY; - Fastweb might re-use some of the aggregated network traffic data it has been collecting for network capacity planning purposes, if useful for processing in PIMCity modules; - LSTech ESPANA will use data collected and used by IMDEA since both are participating in the same task. All the tests will be done by reusing public data, by downloading copies of the target datasets to local development environments and using them in the testbeds;

	- Big Data Analytics takes into account that TE stores transactional Data for the duration of the action. During this period, data remains untouched, and both Data Sellers and Data Buyers can access it to see the history whenever they need to. The TG works similarly, receives raw input data, generates a schema, and serves it to the systems that need that information; - Telefónica Investigación y Desarrollo, Asociación de Usuarios de Internet and CLIQZ do not plan to re-use data; - KU Leuven – CiTiP will re-use the generated research data. (iv) in detail: - Politecnico di Torino will have some data collected by them internally by automatic web crawlers running on clusters of computers. Other will be from public repositories such as https://web.archive.org. Politecnico di Torino plans to continue and enrich the web-crawling archives. The size of the web archives can be very large, up to several terabytes of data depending on the extensiveness of the collection, and the frequencies at which these are collected. The size of the privacy tags archive will be much smaller, in the order of few gigabytes; - NEC Laboratories Europe, Asociación de Usuarios de Internet and CLIQZ will deal with data provided by users; the size of the data to be clarified; - Ermes Cyber Security has collected data internally, using its fleet of automatic web crawlers running on clusters of computers deployed in the cloud; the size of the data is not know yet, i.e. Ermes Cyber Security has collected a few terabytes of data so far, but such size varies depending on a number of variables (e.g, the amount of browsed sites, number of samples, iterations, etc.). The size of the resulting dataset containing D-PMs is expected to be much smaller (few GBs); - Universidad Carlos III de Madrid will handle data from advertising Platforms (closed ones as well as OpenRTB based); the size of the data shall be between ten and thousands of GBs of data; - Telefónica Investigación y Desarrollo will handle data of third party Personal Information Management (PIM) systems (e.g. Facebook, Mobile Phone, Email etc.); the size of the data shall be hundreds of megabytes per user; - the data processed in PIMCity modules hosted in Fastweb's cloud computing infrastructure will come from the sources of said modules, developed and controlled

	by other Consortium parties. The data collected directly by Fastweb will come from network sensors installed in Fastweb's customer network; the size of the data shall be determined later; - IMDEA Networks and LSTech ESPANA public data available in search engines (e.g. Google datasets) or provided by public entities (e.g. https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page). Such datasets are made public without including identifiable personal data for developers and data scientist to test their algorithms; the size of the data strongly depends on the use case, ranging from MB to tens of GB (e.g. mobility data); - Big Data Analytics: the TE stores data from the Offers created by Data Buyers and the transactions that happen when Data Sellers accept selling their data; the TG receives raw input data from the Data Portability and Control tool; with regard to the size of the data, TE stores Offer and transactional data, and expect to generate less than 10GB; the schema generated by the TG occupies less than 1GB; - KU Leuven – CiTiP will generate research data relying on their own expertise in the subject matter; the size of the data will be clarified later. (v) see (i).

2. FAIR DATA	
2.1. Making data findable, including provisions for metadata	Details on (i) whether the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism;[5] (ii) what naming conventions will the partners follow; (iii) whether search keywords will be provided that optimize possibilities for re-use; (iv) whether clear version numbers will be provided; (v) what metadata will be created; in case metadata standards do not exist, details on what type of metadata will be created and how.
	(i) In general, indexes and standards are not relevant in the light of this project; data generally shall not be re-useable; datasets will used for specific goals, however, are not supposed to be useful for general use; (ii) to be defined; (iii) see (i); (iv) public documents will have version numbers; (v) to be defined.
2.2. Making data openly accessible	Details on (i) which data[6] will be made openly available as the default; (ii) whether certain datasets cannot be shared;[7] (iii) how will the data be made accessible;[8] (iv) what methods or software tools are needed to access the data; (v) whether documentation about the software is needed to access the data included; (vi) whether it possible to include the relevant software (e.g. in open source code); (vii) where will the data and associated metadata, documentation and code be deposited;[9] (viii) whether the partners have explored appropriate arrangements with the identified repository; (ix) whether there are restrictions on use, how will access be provided; (x) whether there is a need for a data access committee; (xi) whether there are well-described conditions for access;[10] (xii) how will the identity of the person accessing the data be ascertained.

[5] E.g. persistent and unique identifiers such as Digital Object Identifiers.
[6] Produced and/or used in the project.
[7] Or need to be shared under restrictions. If yes, explanation separating legal and contractual reasons from voluntary restrictions shall be included.
[8] E.g. by deposition in a repository.
[9] Preference should be given to certified repositories which support open access where possible.
[10] I.e. a machine-readable license.

	(i) To be defined, however in any case excluding special categories of data; (ii) to be defined; (iii) in the project website(s); (iv) no specific access tools will be necessary to access deliverables that will be uploaded to the project website(s); software via standard repositories; (v) yes; will be provided; (vi) project partners shall make all software open source as much as possible; will be accessible via the project website(s) and standard repositories; (vii) the project partners will use only internal servers and will not use any public repository or cloud servers; Microsoft Teams will be used for internal storing; the project website(s) – for external access; (viii) not yet; (ix) everything shall be available for free with no restrictions except certain software (to be identified) and certain parts of the documents that may contain intellectual property, trade secrets or other information access to which shall be limited; (x) most likely not since the project does not focus on particularly sensitive data; (xi) see (ix); relevant documents will be provided for accessing software; (xii) people will be able to access most of the deliverables without revealing their identities and personal data of those people will not be stored; for the purposes of statistics (downloads of the deliverables) only aggregated data shall be used (e.g. that certain amount of downloads were in Germany); certain software users will be identified via licencing agreements.

2.3. Making data interoperable	Details on (i) whether the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc.;[11] (ii) what data and metadata vocabularies, standards or methodologies will the partners follow to make the data interoperable; (iii) whether the partners will be using standard vocabularies for all data types present in the data set, to allow inter-disciplinary interoperability; (iv) in case it would be unavoidable that the partners use uncommon or generate project-specific ontologies or vocabularies, will they provide mappings to more commonly used ontologies.
	(i) There are no relevant standards and the partners shall not engage into activities that will focus on creating new standards; various deliverables will be indexed automatically, however; (ii) generally, to be defined, if relevant; with regard to individual partners in detail: - Politecnico di Torino, NEC Laboratories Europe, Ermes Cyber Security, IMDEA Networks, Fastweb, LSTech ESPANA, Big Data Analytics, Asociación de Usuarios de Internet shall define later, if relevant; - N/A to Universidad Carlos III de Madrid, Telefónica Investigación y Desarrollo, CLIQZ and KU Leuven – CiTiP. (iii) no, see (i); (iv) no, see (i).
2.4. Increase data re-use (through clarifying licences)	Details on (i) whether the data will be licensed to permit the widest re-use possible; (ii) when will the data be made available for re-use;[12] (iii) whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project;[13] (iv) how long is it intended that the data remains re-usable; (v) quality assurance processes.

[11] I.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins.

[12] If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

[13] If the re-use of some data is restricted, it shall be explained why.

	(i) Data generally shall not be re-useable; datasets will used for specific goals, however, are not supposed to be useful for general use; only certain software shall be licenced; (ii) within two months after generating particular data; (iii) yes; (iv) five years after the project ends; the project website(s) shall be available for five years after the project ends; (v) N/A.
3. ALLOCATION OF RESOURCES	Details on (i) what are the costs for making data FAIR in the project; (ii) how will these be covered;[14] (iii) who will be responsible for data management in the project; (iv) are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long).
	(i) Costs are mostly represented by two main contributions: - costs to manage and make data accessible for a Fair approach and - costs for hosting data API in servers that hosts data and offer API to access it. (ii) each partner is responsible for the costs incurred in making data FAIR, and for open access; (iii) responsible partners are indicated in the table at the beginning of this document. In particular, specific roles are foreseen for technical coordinator (T8.2); dissemination manager (T6.1); innovation manager (T6.4); data manager (T7.1). Besides, PIMCity data protection officer team, subject to their competences, will advise the PIMCity project partners on data management; (iv) to be defined.
4. DATA SECURITY	Details on (i) what provisions are in place for data security;[15] (ii) whether the data safely stored in certified repositories for long term preservation and curation.
	(i) Each partner adopts individual approach; to be clarified later; (ii) no since we do not use repositories; specifics of the project have to be taken into account.

[14] Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).
[15] Including data recovery as well as secure storage and transfer of sensitive data.

5. ETHICAL ASPECTS	Details on (i) whether any ethical or legal issues can have an impact on data sharing;[16] (ii) whether the informed consent for data sharing and long-term preservation is included in questionnaires dealing with personal data.
	(i) Defined in the deliverables of WP7 and WP9 (in the deliverables coordinated by KU Leuven – CiTiP and POLITO) and going to be refined at the later stages of the project; (ii) the project partners will comply with all of the relevant personal data protection requirements.
6. OTHER	Details on whether there is any use of other national/funder/sectorial/departmental procedures for data management and, If yes, references to the particular ones.
	The project partners are following best industry practices and operate in compliance with internal procedures. In case the project partners would agree to follow any particular procedures, it will be provided in the updated version(s) of the data management plan.

ANNEX A – DATA PROTECTION IMPACT ASSESSMENT

Report of the initial result of the Data Protection Impact Assessments carried out in the PIMCity's project context

Legal Basis

According to Art. 35 of the Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data[17] (hereinafter as the GDPR), a data protection impact assessment (hereinafter as a DPIA) should be carried out when the processing

[16] These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
[17] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, 4.5.2016, p. 1–88.

activity of personal data is *'likely to result in a high risk for the rights and freedoms of natural persons'* (Art. 35(1) GDPR).[18] The methodology used to develop the DPIA is based not only the GDPR provisions, but also the guidelines provided by Article 29 Working Party (hereinafter as WP29).[19] To this end, WP29, substantiating GDPR provision lists 'nine criteria' that should be taken into account to establish whether a DPIA should be carried out.[20] The activities characterising the PIMCity project are likely to fall into the two listed criteria. In particular: *'Data processed on a large scale'*[21] and *'Matching or combining datasets'*.[22]

In line with the legal requirements foreseen under Art. 35 GDPR, and in compliance with the privacy and data protection accountability principle,[23] PIMCity partners have carried out the DPIA in the light of the PIMCity project.

Based on Art. 35(7) GDPR, a DPIA shall include:

- a description of the context and purposes of the processing of personal data (context);
- a justification of the necessity and proportionality of the processing operations in relation to the purposes (fundamental rights);
- an assessment of the risks to the rights and freedoms of data subjects that might be generated by the processing activities;
- an explanation of the measures envisaged to address the risks, including safeguards, security measures, and mechanisms to ensure the protection of personal data and to demonstrate compliance with the GDPR (mitigation measures).

[18] Art.35(1) GDPR: *'Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks'*.

[19] WP29, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP248 rev.01, adopted on 4 April 2017 as last revised and adopted on 4 October 2017.

[20] Ibid, p.10

[21] Art. 35(3)(b) GDPR. Large scale of data processing is included among the examples that GDPR provides that require a DPIA. Nonetheless, the GDPR does not provide a definition of what constitutes large-scale, though recital 91 provides some guidance. Contrary, according to WP29 guidelines following factors should be considered to determine whether the processing is carried out on a large scale: *'a. the number of data subjects concerned, either as a specific number or as a proportion of the relevant population; b. the volume of data and/or the range of different data items being processed; c. the duration, or permanence, of the data processing activity; d. the geographical extent of the processing activity'*. WP29, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP248 rev.01, adopted on 4 April 2017 as last revised and adopted on 4 October 2017, p.10.

[22] Matching or combining criteria concerns for example processing activities originating from two or more data processing operations performed for different purposes and/or by different data controllers in a way that would exceed the reasonable expectations of the data subject. Such criteria is strictly related to the purpose limitation principle, therefore, specific attention should be paid in regard to the lawful basis for processing.

[23] Art. 5(2) GDPR: *'The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability').'*

Methodology

To achieve comparable results, PIMCity partners have opted for a common approach concerning tools and methodology for the execution of the DPIA.

First, KU Leuven – CiTiP has provided clarifications and guidelines on the DPIA, and how this should be developed. Consortium partners NEC Laboratories Europe, Politecnico di Torino, Fastweb, Asociación de Usuarios de Internet, Universidad Carlos III de Madrid, Telefónica Investigación y Desarrollo, IMDEA Networks, CLIQZ (hereinafter as the PIMCity partners) have concluded that, given their roles and foreseen activities in the PIMCity project, they shall carry out a DPIA.

Second, partners whose activity fall into the GDPR scope of application, have carried out its respective DPIAs individually. Following the suggestion of KU Leuven – CiTiP, the partners have developed their DPIA through the software PIA, provided by the French Data Protection Authority 'CNIL' ('Commission Nationale de l'Informatique et des Libertés').[24] CNIL software has been chosen for two main reasons: it is developed by a national data protection authority, and include all requirements listed in Art. 35(7) GDPR.[25]

Third, the final results of the PIMCity partners individual DPIA have been shared within the consortium through the project repository made available in Microsoft 'Teams'.

Finally, the results provided by partners have been collected and summarised by KU Leuven – CiTiP and provided in the table below.

More details on various mitigation measures are provided in D9.2 (confidential), coordinated by Politecnico di Torino.

[24] https://www.cnil.fr/en/privacy-impact-assessment-pia
[25] Erik Kamenjasevic, Elisabetta Biasin, Safecare - Deliverable 6.1 Legal and ethical inventory and in-depth analysis, December 2018

Initial PIMCity Data Protection Impact Assessment

DPIA Sections	Task	Partners' Initial Assessment
CONTEXT	Overview of the processing activities and purposes under consideration	PIMCity partners have described the activities involving the processing of personal data and the purpose of such activities. They have also described the nature of the data processed. Sample activities of PIMCity partners include data collection, data storage, data use, data combination, and erasure. PIMCity partners have also described and how they intend to achieve the declared purpose (e.g. collects and processes data from websites using automatic web scrapers to generate privacy metrics).
FUNDAMENTAL RIGHTS	An evaluation of the necessity and proportionality of the processing operations in relation to the purposes	PIMCity partners, according to their role within the project have declared that the processing purpose of their activity is explicit and legitimate. Concerning lawful basis for processing personal data, most of the PIMCity partners have affirmed that consent is going to be the legal ground used for their processing activities. Besides, PIMCity partners have also provided information about their activities regarding data accuracy, data minimisation, and data storage.
RISK ASSESSMENT	Assessment of the risks to the rights and freedoms of data subjects resulting from the processing activities	PIMCity partners have assessed the risks to the rights and freedoms of data subjects link to their activities. According to PIMCity partners risks can be generated by: (I) Major Event (II) System bug (III) Third-party unauthorised access (IV) Human error Most of the assessments show a negligible risk. A minor part of these assessments find a limited risk.
MITIGATION MEASURES	The mitigation measures envisaged for addressing the risks[26]	PIMCity partners have already developed a list of security measures and protocols to address the highlighted risks, taking into account the state-of-the-art of such security protocols. In particular, PIMCity partners mitigation measures include: (I) Physical access control, (II) Logical access control, (III) Hardware access control. In particular, the security of personal data partners will be ensured by encryption protocols and when necessary anonymisation functions.

[26] Erik Kamenjasevic, Elisabetta Biasin, *Safecare - Deliverable 6.1 Legal and ethical inventory and in-depth analysis*, December 2018.

ANNEX B – GUIDELINES FOR CONSENT MANAGEMENT

The guidelines provide non-exhaustive recommendations for consent management, taking into account the requirements provided by the Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data[27] (hereinafter as the GDPR) and the recommendations provided by the European Data Protection Board (hereinafter as the EDPB).[28] Template informed consent form is provided as Annex D and shall be adjusted (tailored) by the partners' on a case by case basis taking into account particular details.

Consent is a concept used both in the GDPR and the Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (hereinafter as the e-Privacy Directive)[29] and is one of the legal grounds for personal data processing that may be given for one or more specific purposes.[30]

First, it is important that consent shall always be obtained before the controller starts processing personal data for which consent is needed.[31] Second, Art. 4(11) and Rec. 32 of the GDPR provides that consent shall be given '**by a clear affirmative act**' establishing a '**freely given**', '**specific'**, '**informed'** and '**unambiguous'** indication of the data subject's agreement to the processing of personal data. Besides the Art. 4(11) and Rec. 32 the GDPR provides requirements in Rec. 33, 42 and 43 as to how the controller must act to comply with the main elements of the consent requirement.[32] Each and every requirement shall be fulfilled, taking into account the relevant case law of the Court of Justice of the European Union and the recommendations (guidelines) provided by the competent institutions such as the EDPB or, if relevant, certain national data protection authorities. Generally, consent can only be an appropriate lawful basis if a data subject is offered control and is offered a genuine choice with regard to accepting or declining the terms offered or declining them without detriment.[33]

Freely given

For the consent to be freely given the data subject shall be able to genuinely exercise its autonomy. As the EDPB puts it, the element 'free' implies 'implies real choice and control

[27] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, 4.5.2016, p. 1–88.
[28] EDPB, Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.1, adopted on 4 May 2020, available at <https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf>, accessed 14/05/2020.
[29] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] OJ L 201, 31.7.2002, p. 37–47. The GDPR conditions for obtaining valid consent are applicable in situations falling within the scope of the e-Privacy Directive. EDPB, Guidelines 05/2020 on consent, p. 5.
[30] Art. 6(1) GDPR.
[31] EDPB, Guidelines 05/2020 on consent, p. 18.
[32] Ibid, p. 5.
[33] Ibid, p. 4.

for data subjects'.[34] In relation to this, certain recommendations shall be taken into account, including but not limited:

- consent cannot be bundled up as a non-negotiable part of terms and conditions, e.g. consent for marketing cannot be hidden in the general terms and conditions of the website;
- there cannot be any element of pressure or influence, e.g. data subject cannot be required to consent for marketing in order to use basic features of an app;
- data controllers shall be particularly careful while relying on consent if they act as public authorities or employers since it is considered that there is often a clear imbalance of power in such relationships, hence consent may be regarded as not freely given; it is recommended to consider alternative legal basis for processing personal data;
- data controllers shall avoid 'bundling' consent with acceptance of terms or conditions, or 'tying' the provision of a contract or a service to a request for consent to process personal data that are not necessary for the performance of that contract or service.[35]

Cookie walls

The EDPB has clarified that access to services and functionalities 'must not be made conditional on the consent of a user to the storing of information, or gaining of access to information already stored, in the terminal equipment of a user (so called cookie walls).'[36] Hence, it is not allowed to, e.g. block the access to the website unless the data subject accepts cookies.

Specific

The consent shall be provided for specific, particularly clear purposes (for each purpose separately) and the information regarding the consent shall be provided separately from all of the other information. As EDPB puts it, data subjects shall give their consent with the understanding that they are in control and their data will only be processed for those specified purposes.[37] E.g. requesting consent for 'business purposes' would not be specific enough.

Informed

In essence, it is considered that the consent cannot be considered meaningful in case data subject is not informed about the relevant details properly. In relation to this, a number of details shall be provided. The EDPB is of the opinion that at least the following information is required for obtaining valid consent:

- data controller's (or multiple (joint) controllers') identity(-ies);
- the purpose of each of the processing operations for which consent is sought;
- what (type of) data will be collected and used;
- the existence of the right to withdraw consent;

[34] Ibid, p. 6.
[35] For more details see EDPB, Guidelines 05/2020 on consent, p. 6-11.
[36] EDPB, Guidelines 05/2020 on consent, p. 11.
[37] Ibid, p. 13.

- information about the use of the data for automated decision-making[38] where relevant, and
- the possible risks of data transfers due to absence of an adequacy decision and of appropriate safeguards.[39]

Data controller is not obliged to provide a list of particular data processors, however shall provide a list of recipients or categories of recipients.[40]

The information regarding consent shall be provided separately from all of the other information such as terms and conditions, in a very clear and plain language.[41] Information regarding consent can be provided in written or oral statements, as well as in audio or video messages[42].

Unambiguous – form of consent

Data controller shall be able to demonstrate that it is obvious that data subject consented to the particular processing of personal data. While it is rather self-explanatory when it comes to the written agreements which require signature, there are some grey zones in the electronic environment, discussed in detail below.

Consent may be given not only as a written statement but also by electronic means. This could include 'ticking a box when visiting an internet website, choosing technical settings for information society services or another statement or conduct which clearly indicates in this context the data subject's acceptance of the proposed processing of his or her personal data. Silence, pre-ticked boxes, opt-out constructions, inactivity as well as merely proceeding with a service do not constitute consent'.[43] In case consent shall be given following a request by electronic means, 'the request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided.'[44] In other words, it is recommended to interrupt the user experience to ensure there is no ambiguity.

Nevertheless, the above is not to suggest that ticking the box is the only option. The users may be suggested to swipe or to perform other actions as long as sufficient information is provided and these actions would be sufficient to signify an agreement to a particular request. However, 'scrolling or swiping through a webpage or similar user activity will not under any circumstances satisfy the requirement of a clear and affirmative action: such actions may be difficult to distinguish from other activity or interaction by a user'.[45]

Besides, consent may be given as an oral statement. However, in all of the cases while choosing the particular form data controller shall take into account that it shall be able to demonstrate that consent was given (with regard to the latter see also the paragraph on *demonstration* below).

[38] In accordance with Art. 22 (2)(c) of the GDPR.
[39] As described in Art. 46 of the GDPR. EDPB, Guidelines 05/2020 on consent, p. 14.
[40] Ibid.
[41] Ibid, p. 16.
[42] Ibid.
[43] Rec. 32 GDPR; EDPB, Guidelines 05/2020 on consent, p. 17.
[44] Ibid.
[45] Ibid, p. 18.

Explicit consent

Data controller shall take into account that *explicit* consent is required in certain cases such as for processing of special categories of data or for automated individual decision-making, including profiling.[46] Data controller shall consider requiring signed written statement, filling of an electronic form, receiving an email, receiving an uploaded scanned document carrying the signature or using electronic signature.[47] EDPB also suggests to consider two stage verification involving communication via email and SMS messages.[48]

Demonstration and withdrawal of consent

For consent to be valid the controller shall be able to demonstrate that consent was given as well as to provide subject with the opportunity to withdraw it since obtaining consent should be a reversible decision[49].

Demonstration

Data controller is free to choose its own way of demonstrating that consent was given, however it shall prove not only, e.g. the affirmative action, but also that all of the conditions were fulfilled, e.g. that consent was informed, freely given, etc. After the end of the processing activity, such proof should be kept no longer then strictly necessary for compliance with a legal obligation or for the establishment, exercise or defence of legal claims.[50]

Withdrawal

Data controller shall ensure data subject can withdraw the consent any time. It is not required that the consent would be withdrawn in the same way as it was given, however it is important that this procedure is not complicated and certainly not requiring more effort than for giving the consent. E.g. if the consent was given through one mouse click, the same amount of effort shall be sufficient to withdraw (e.g. requesting to send an e-mail, let alone a registered mail would be excessive). Besides, data subject shall be able to withdraw free of charge and shall face no negative effects with regard to services.[51]

Regarding the purposes of data processing – granularity requirements

Consent 'should cover all processing activities carried out for the same purpose or purposes. When the processing has multiple purposes, consent should be given for all of them'.[52]

In particular, in case personal data will be processed for multiple purposes, data subjects shall be free to choose with regard to each purpose separately. For example, it cannot be required to consent to processing for marketing and analytical purposes with one tick.

[46] For more details see Guidelines 05/2020 on consent, p. 19.
[47] Ibid.
[48] For particular guidelines see EDPB, Guidelines 05/2020 on consent, p. 19.
[49] Ibid, p. 5.
[50] In accordance with Art. 17(3)(b) and (e) GDPR. For more details see EDPB, Guidelines 05/2020 on consent, p. 21.
[51] For more details see EDPB, Guidelines 05/2020 on consent, p. 22.
[52] Rec. 32 GDPR.

***

Finally, obtaining consent does not negate or in any way diminish the controller's obligations to observe the principles of processing enshrined in the GDPR, 'especially Article 5 of the GDPR with regard to fairness, necessity and proportionality, as well as data quality'.[53] Although there is no specific time limit with regard to how long the consent may last, the EDPB recommends data controllers to refresh it periodically.[54]

For more details, including practical examples and details on specific areas of concern in the GDPR such as children and scientific research see the guidelines of the EDPB.[55] For the additional recommendations regarding consent management on the project's website(s) please also see Annex B of the D7.2.

Data controller shall also ensure compliance with the relevant national legal requirements.

ANNEX C – GUIDELINES FOR PRIVACY POLICIES

The guidelines provide non-exhaustive recommendations for privacy policies, taking into account the requirements provided by the Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data[56] (hereinafter as the GDPR). Template privacy policy is provided as Annex E and shall be adjusted (tailored) by the partners' on a case by case basis taking into account particular details.

The GDPR provides an obligation for the data controller to take into account the 'nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons'[57], and to implement 'appropriate technical and organisational measures' to ensure and to be able to demonstrate that processing is performed in accordance with the GDPR accordingly. Where proportionate in relation to processing activities, these measures shall include the implementation of appropriate data protection policies.[58] In relation to this, there is a number of circumstances where privacy policies shall be considered to be necessary.

Content

In principle, a privacy policy shall provide all of the essential details related to data processing such as the categories of personal data being processed, processing purposes, third parties that may get access to personal data, etc. While considering which particular information shall be included in a privacy policy, certain provisions of the GDPR shall be taken into account.[59]

[53] EDPB, Guidelines 05/2020 on consent, p. 4.
[54] Ibid, p. 22.
[55] EDPB, Guidelines 05/2020 on consent.
[56] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, 4.5.2016, p. 1–88.
[57] Art. 24(1) GDPR.
[58] Art. 24(1) and 24(2) GDPR.
[59] Including but not limited Art. 12-15 GDPR.

In the light of the PIMCity project it is relevant that personal data will be collected from the data subject. In such a case privacy policy shall be accessible no later than at the time when personal data are obtained, and shall provide these details:

1. categories of personal data collected (to be processed);[60]
2. controller's (and its representative's, if applicable) **identity and contact details**;[61]
3. **contact details of the data protection officer**, if applicable;
4. **purposes** of personal data processing**;**
5. **legal basis(-es)** for personal data processing;
6. **legitimate interests** of the controller or a third party, in case processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party;
7. **recipients** or categories of recipients of personal data, if any;
8. where applicable, **the fact** that the controller intends to transfer personal data to a third country or international organisation **and the existence or absence of an adequacy decision** by the European Commission, **or in the case the transfers would be based on appropriate safeguards or binding corporate rules,[62] or on the second subparagraph of Art. 49(1) of the GDPR, reference to the appropriate or suitable safeguards** and the means by which to obtain a copy of them or where they have been made available;
9. **period** for which personal data will be stored;[63]
10. reference to the right to request access to and rectification or erasure of personal data or restriction of processing or to object to processing as well as the right to data portability;
11. in case processing is based on **consent**, **right to withdraw** it;
12. the right to lodge a complaint with a supervisory authority;
13. whether the provision of personal data is (i) a **statutory** or **contractual requirement**, **or** (ii) a requirement **necessary to enter into a contract**, as well as (iii) whether the data subject is **obliged** to provide personal data and (iv) of the possible **consequences of failure to provide** such data;
14. the existence of **automated decision-making**, including **profiling**, referred to in the Art. 22(1) and (4) of the GDPR **and**, at least in those cases, **meaningful information about the logic involved**,[64] as well as the **significance and the envisaged consequences of such processing** for the data subject.[65]

In case the controller would like to process personal data for a purpose other than that for which the personal data were collected, the controller shall provide the data subject prior to that further processing with information on that other purpose and with any relevant information identified in the lines 9-14 above.[66]

[60] E.g. names, surnames, IP addresses, etc.
[61] The requirements listed in lines 2-14 are stemming from Art. 13 GDPR.
[62] As referred to in Art. 46 or 47 GDPR.
[63] In case it is not possible, the criteria used to determine that period shall be provided.
[64] With regard to alternative approaches to providing such information see, e.g. M. Brkan and G. Bonnet, 'Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas', 11 (2020) *European Journal of Risk Regulation*, https://doi.org/10.1017/err.2020.10.
[65] Art. 13 GDPR.
[66] Art. 13(3) GDPR.

Where and insofar as data subject already has the information listed above, it is not necessary to provide it additionally.[67]

In case personal data have not been obtained from the data subject, the controller shall provide the data subject with the information required by the Art. 14 of the GDPR.[68]

Form

Information provided in a privacy policy shall be concise, transparent, intelligible, easily accessible, written in clear and plain language, particularly if addressed to a child, and free of charge.[69] The information may be provided in combination with standardised icons in order to give in an easily visible, intelligible and clearly legible manner a meaningful overview of the intended processing. Where the icons are presented electronically, they shall be machine-readable.[70]

For the additional recommendations regarding cookies management on the project's website(s) please also see Annex B of the D7.2.

***

Data controller shall also ensure compliance with the relevant national legal requirements.

ANNEX D – TEMPLATE INFORMED CONSENT FORM[71]

[Name of the Data Controller]

Consent for Processing of Personal Data

This document provides you an opportunity to allow us to process your personal data and information that shall help you to decide whether you want to allow that.[72] In particular, this document provides details about (i) who would be controlling your data; (ii) for what purposes your data would be processed; (iii) what (type of) your data would be processed; and reminds that you always have a right to withdraw your consent.

[67] Art. 13(5) GDPR.
[68] Art. 14 GDPR.
[69] Art. 12 GDPR.
[70] Art. 12(7) GDPR.
[71] [The template consent form shall be adjusted (tailored) by the partners' on a case by case basis taking into account particular details. This template consent form is drafted for the cases where there is no automated decision-making and no risks related to absence of an adequacy decision and of appropriate safeguards.]
[72] As required by the Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, 4.5.2016, p. 1–88, Art. 4(11), Rec. 32, and recommended by the European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.1, adopted on 4 May 2020, p. 14.

Your personal data would be processed by [Name of the data controller], a [Company type], with registered addressed at [Address] (hereinafter as 'Controller' or 'We').

Your personal data, in particular **[Provide categories of personal data]** would be processed **for [Description of the purposes in detail].**

In case you would have any **questions** or **concerns** with regard to processing of your data, you can always contact the us by sending an e-mail [E-mail address of the data protection officer (team) or other responsible person]. You could also **withdraw your consent at any time** by sending an e-mail to the same e-mail address.

Detailed information about your rights and processing of your data could be found at [Link to a website, e.g. privacy policy].

In case you agree that we would process your data under the conditions described above, please kindly sign this document:

[In case there are several purposes, there shall be an opportunity to express will with regard to every purpose separately. In case consent would be requested by electronic means, data controller shall ensure all of the relevant information is provided, revealed in detail in Annex B].

_____

Name:
Date:

ANNEX E – TEMPLATE PRIVACY POLICY[73]

[Name of the Data Controller]

Privacy Policy

Last revised on [Date]

Please read this privacy policy carefully so that you fully understand how we collect, use and store information about you. In case you accept this privacy policy, we assume that you understand and agree with all the details provided below. Occasionally, we may update this privacy policy, so please kindly visit our website to always find the latest version. We will contact you in case there will be important changes.

[73] [The template privacy policy shall be adjusted (tailored) by the partners' on a case by case basis taking into account particular details. This template privacy policy is drafted for the cases where personal data are collected from the data subject. It shall be accessible at the time when personal data are obtained.]

1. Introduction

1.1. [Name of the data controller], a [Company type], with registered addressed at [Address] (hereinafter as 'the Controller' or 'we') protects your information and your privacy and processes your personal data following the requirements provided by the relevant laws.[74]

1.2. This privacy policy (hereinafter as 'the Privacy Policy') describes how we process your personal data. In particular, this Privacy Policy provides details on [Please review and adjust the section below carefully; please only leave those sections which are relevant, e.g. consider deleting sections 1.2.3, 1.2.6, 1.2.7, 1.2.8, 1.2.11 and/or 1.2.14 in case they appear irrelevant; remove unnecessary details such as phrases 'if applicable' and 'if any']:

1.2.1. personal data collected (to be processed);[75]
1.2.2. controller's (and its representative's, if applicable) **identity and contact details**;[76]
1.2.3. **contact details of the data protection officer**, if applicable;
1.2.4. **purposes** of personal data processing**;**
1.2.5. **legal basis(-es)** for personal data processing;
1.2.6. **legitimate interests** of the controller or a third party, in case processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party;
1.2.7. **recipients** or categories of recipients of personal data, if any;
1.2.8. where applicable, **the fact** that the controller intends to transfer personal data to a third country or international organisation **and the existence or absence of an adequacy decision** by the European Commission, **or in the case the transfers would be based on appropriate safeguards or binding corporate rules,[77] or on the second subparagraph of Art. 49(1)** of the Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data **(hereinafter as the GDPR), reference to the appropriate or suitable safeguards** and the means by which to obtain a copy of them or where they have been made available;
1.2.9. **period** for which personal data will be stored [in case it is not possible, provide the criteria used to determine that period shall be provided];
1.2.10. reference to the right to request access to and rectification or erasure of personal data or restriction of processing or to object to processing as well as the right to data portability;
1.2.11. in case processing is based on **consent**, **right to withdraw** it;
1.2.12. the right to lodge a complaint with a supervisory authority;
1.2.13. whether the provision of personal data is (i) a **statutory** or **contractual requirement**, **or** (ii) a requirement **necessary to enter into a contract**, as well as (iii) whether the data subject is **obliged** to provide personal data and (iv) of the possible **consequences of failure to provide** such data;

[74] As required by the Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, 4.5.2016, p. 1–88.
[75] E.g. names, surnames, IP addresses, etc.
[76] The requirements listed in 1.2.2-1.2.14 are stemming from Art. 13 GDPR.
[77] As referred to in Art. 46 or 47 GDPR.

1.2.14. the existence of **automated decision-making**, including **profiling**, referred to in the Art. 22(1) and (4) of the GDPR **and**, at least in those cases, **meaningful information about the logic involved**,[78] as well as the **significance and the envisaged consequences of such processing** for the data subject.[79]

[To ensure that information is provided in an easily accessible manner, consider providing certain information in a scheme, e.g. in a table. For example, consider providing the information of section 2 and section 3 in a table of three columns and multiple lines, where the first column would provide inforamtion about the purposes of processing, the second – inforamtion about the particular data being processed for the respective purpose, the third – legal basis for processing of that particular data and reference to the GDPR].

2. Which information do we collect?

2.1. We collect and process [Please provide particular details with regard to personal data which is collected and will be processed, e.g. to inform data subject – name, surname, e-mail address; to handle queries – query, request or complaint, information, related to the query, request or complaint, communication with the Controller; to engage in legal proceedings relating to data subject – all of the information mentioned above, documents and attachments submitted by you, procedural documents and court documents; including but not limited].

3. Why do we collect information about you?

3.1. [Please provide clearly and separately in a form of list the particular purposes of personal data processing, e.g. performing data analyses (including anonymization and aggregation of personal data), creating individual profiles and using and sharing the resulting data to third parties for commercialisation or research purposes; preventing illegal activities; complying with and enforcing any applicable laws].

3.2. [Please provide legal basis(-es) for personal data processing clearly and separately in a form of list. Please be kindly reminded that legal bases include (a) consent (with reference to Art. 6(1)(a) of the GDPR); (b) necessity for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract (with reference to Art. 6(1)(b) of the GDPR); (c) necessity for compliance with a legal obligation to which the controller is subject (with reference to Art. 6(1)(c) of the GDPR); (d) to protect the vital interests of the data subject or of another natural person (with reference to Art. 6(1)(d) of the GDPR) [unlikely to be relevant]; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller (with reference to Art. 6(1)(e) of the GDPR); (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject

[78] With regard to alternative approaches to providing such information see, e.g. M. Brkan and G. Bonnet, 'Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas', 11 (2020) *European Journal of Risk Regulation*, https://doi.org/10.1017/err.2020.10.

[79] Art. 13 GDPR.

which require protection of personal data, in particular where the data subject is a child; in the latter case (of legitimate interests, please clarify them in detail[80]] (with reference to Art. 6(1)(f) of the GDPR). Please note the latter may cover processing queries, requests and complaints submitted by data subject for the purposes of handling them, as well as processing of various documents submitted by data subject or concerning data subject for the purposes of engaging to legal proceedings, including but not limited.

4. Which information do you have to provide and why?

4.1. You should provide us with the information that is necessary to handle your complaints, queries and requests. In case you would not provide us with such information, we would not be able to handle your complaints, queries and requests.

5. Is your information shared with anyone else?

5.1. [Please identify the recipients or categories of recipients of personal data.]

5.2. We may also share your personal data to courts and other institutions or subjects when it is required by law.

5.3. Other than as set out in this Privacy Policy, we shall not disclose your personal data to any third parties without obtaining your prior explicit consent unless that would be required by law.

6. Is your information transferred to a third country or international organisation?

6.1. [Please provide information regarding personal data transfers to a third country(-ies) and/or international orgnisation(s), including information indicated in 1.2.8; alternatively remove this section].

7. For how long do we keep your information?

7.1. [Please provide information on how long personal data will be stored [in case it is not possible, provide the criteria used to determine that period shall be provided].

8. How do we secure your information?

8.1. [Please provide information about the security measures].

8.2. The Controller shall take appropriate administrative, technical and organizational measures against unauthorized or unlawful processing of any personal data or its accidental loss, destruction or damage, access, disclosure or use.

8.3. In the event of and following discovery or notification of a breach of the security of the personal data, or access by an unauthorised person, the Controller shall notify you if the breach is likely to affect your privacy.

[80] Art. 6(1) GDPR.

9. How do we manage cookies? [Remove if not applicable]

[To ensure that information is provided in an easily accessible manner, consider providing certain information in a scheme, e.g. in a table.]

9.1. [Please provide information about the cookies – (i) cookie name, (ii) cookie category, (iii) cookie purpose and (iv) cookie expiry, each respectively].

9.2. We will not use cookies for which your consent is necessary without your consent.

9.3. Please note that you can configure your browser to decline all or some cookies as well as to ask your permission for accepting them.

9.4. To understand and control cookies of the other companies or institutions (so-called third-party cookies), please read the policies of such third-parties.

10. What are your rights?

10.1. You can ask the Controller whether it processes the data about you and, if yes, you can request access to that data**.**
10.2. You can ask the Controller to correct the inaccurate data about you.
10.3. You can ask to Controller to delete the data about you and the data shall be deteled in case there are no legal basis for the Controller to process it.
10.4. You can ask the Controller to restrict processing of your data (i) in case you contest the accuracy of the data, (ii) in case the processing is unlawful and you oppose the erasure of it, (iii) in case the Controller no longer needs the personal data but they are required by the data subject for the legal claims, or (iv) in case you have objected to processing pursuant to Art. 21(1) of the GDPR pending the verification whether the legitimate grounds of the controller override those of the data subject.
10.5. You can object processing your data in case it is processed for third party's or Controller's interests.
10.6. You can ask to transfer (receive) your data which is processed automatically and is provided to us by your consent or under the contract.
10.7. You can withdraw your consent given to us regarding processing of your personal data.
10.8. You have the right to request us not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning you or similarly significantly affects you [if applicable].
10.9. To complain to a supervisory authority and to seek a judicial remedy.

[Name of the relevant supervisory authority]
[Address of the relevant supervisory authority]
[Contacts of the relevant supervisory authority]

How can you contact us?

Data Controller
[Name of the data controller], a [Company type], with registered addressed at [Address]
[Contact details]

Data Protection Officer, if applicable
[Name] [Contact details]

ANNEX F – PARTNERS' INDIVIDUAL INPUT FOR THE DATA MANAGEMENT PLAN

1. DATA SUMMARY

1.1. Politecnico di Torino

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The purpose is to devise and test algorithms to automatically define privacy tags. Privacy tags refers to website and webservices, and not to individuals. This is one of the specific goals defined in WP2.

What types and formats of data will the project generate/collect?

We will collect information instrumental to check if and how the website or web service collects and exchanges eventual personal information. Privacy tags will summarize the output of the algorithms.

Will you re-use any existing data and how?

We have historical web crawling archives that have been performed in the past. These contains a snapshot of web pages done through the time using automatic web crawlers.

What is the origin of the data?

Some have been collected by POLITO internally by automatic web crawlers running on clusters of computers. Other are coming from public repositories such as https://web.archive.org. We plan to continue and enrich the web-crawling archives.

What is the expected size of the data?

The size of the web archives can be very large, up to several terabytes of data depending on the extensiveness of the collection, and the frequencies at which these are collected. The size of the privacy tags archive will be much smaller, in the order of few gigabytes.

To whom might it be useful ('data utility')?

These web archives are commonly used for running algorithms to detected privacy violations from website and webservices.
These are also useful to check website performance (e.g., web page load time), and possible issues with pages (e.g., errors for missing objects).

Privacy tags will be useful for the project goals and for anyone interested in the web data economy.

1.2. NEC Laboratories Europe

What is the purpose of the data collection/generation and its relation to the objectives of the project?

NEC will collect the hosts visited by users to profile their interests. It is one of the main tasks parts of T4.4 and key of the personal data avatar.

What types and formats of data will the project generate/collect?

The data will be composed by sequences of host, it will be stored in a mysql database.

Will you re-use any existing data and how?

Probably we will re-use anonimyzed dataset containing traffic logs.

What is the origin of the data?

The users. To be clarified.

What is the expected size of the data?

In the order of Gb of data. To be clarified.

To whom might it be useful ('data utility')?

To us. To be clarified.

1.3. Ermes Cyber Security

What is the purpose of the data collection/generation and its relation to the objectives of the project?

ECS will design, develop and test algorithms to automatically generate Privacy Metrics (D-PM) which will be ultimately used to generate Transparenct Tags. Transparency Tags describe which personal data is collected by websites and web services (e.g., browsing history). As such no individual will be involved and no personal data will be collected in this task. This is one of the specific goals defined in WP2.

What types and formats of data will the project generate/collect?

ECS will collect data generated by its fleet of automatic web scrapers, which will be then processed by automatic algorithms to identify which personal data web services collect and how. The result of this analysis, the D-PM, will be provided in JSON format, and stored and distributed using state-of-the-art database technologies.

Will you re-use any existing data and how?

ECS will re-use data collected in the past using its fleet of automatic web scrapers. This contains information about web sites including the code used to generate the page (e.g., HTML and Javascript) as well as logs describing the APIs executed by the browser to generate the page.

What is the origin of the data?

ECS has collected such data internally, using its fleet of automatic web crawlers running on clusters of computers deployed in the cloud.

What is the expected size of the data?

ECS collected few terabytes of data so far, but such size varies depending on a number of variables (e.g, the amount of browsed sites, number of samples, iterations, etc.). The size of the resulting dataset containing D-PMs is expected to be much smaller (few GBs).

To whom might it be useful ('data utility')?

Apart from the generation of D-PMs, collected data is used by ECS to feed several algorithms with different purposes such as website classification and vulnerability assessment.

D-PMs are the fundamental brick at the base of Transparency Tags and will be used by the partners to achieve the project goals, and by anyone interested in understanding the usage of data in the web.

1.4. IMDEA Networks

What is the purpose of the data collection/generation and its relation to the objectives of the project?

IMDEA will require data to test and demonstrate the functionality of data valuation tools for combinations of sources from the users' perspective and the proposed data marketplace concepts provided. The objective of such processing will be to find the response of an AI/ML algorithm to different combinations of inputs for a set of available sources or users, to calculate the value they are bringing to the specific task.
IMDEA will also develop methodologies to modify the traded dataset and add some kind of watermark to the dataset that eventually helps track potential leakages of information by trusted partners. Should a copy of that information be made public in the Internet, the watermark should help identify the entity that shared it.

What types and formats of data will the project generate/collect?

IMDEA will leverage existing public data available on the Internet that may eventually resemble the kind of information that the PDK or the EasyPIMS would be managing and trading. This is in general structured data, for instance anonymized mobility data or CDRs (call detail records) showing the mobility of people within an area, sich as a city. More details could be provided once the specific use cases to implement are clear. As of now, we will be working with this mobility information which in all cases is totally anonymized.

Will you re-use any existing data and how?

All the tests will be done by reusing public data, by downloading copies of the target datasets to local development environments and using them in the testbeds.

What is the origin of the data?

Public data available in search engines (e.g. Google datasets) or provided by public entities (e.g. https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page). Such datasets are made public without including identifiable personal data for developers and data scientist to test their algorithms.

What is the expected size of the data?

It will strongly depend on the use case, ranging from MB to tens of GB (e.g. mobility data).

To whom might it be useful ('data utility')?

Mobility data is useful for city planners or transportation enterprises at the time of planning resources or operations in their enterprises. CDR from mobile operators are being actively used in the billing process or for network planning purposes. Governments are using anonymized mobility information to track mobility in the city, for instance recently to measure the degree of mobility of the population during the Coronavirus outbreak.

1.5. Universidad Carlos III de Madrid

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The purpose is to identify the value that market assigns to end-users' value. This is one of the specific goals defined in WP3.

What types and formats of data will the project generate/collect?

We will collect data regarding the value/price of audiences (i.e., user profiles) in online advertising platforms.

Will you re-use any existing data and how?

We have some data collected before the start of the project, in the context of TYPES and SMOOTH EU projects and we might reuse it. It is of the same nature and type than the one we plan to collect in PIMCITY.

What is the origin of the data?

Advertising Platforms (Closed ones as well as OpenRTB based).

What is the expected size of the data?

Between ten and thousands of GBs of data.

To whom might it be useful ('data utility')?

To PIMs in order to handle the offer of data on behalf of their user base. Third parties willing to bid (or request the use) of data in exchange of a monetary transaction.

1.6. Telefónica Investigación y Desarrollo

What is the purpose of the data collection/generation and its relation to the objectives of the project?

We collect and process data to enable data migration to new platforms in a privacy-preserving fashion. More specifically users will be able to download their data from other Personal Information Management (PIM) systems (e.g. Mobile Phone, Email etc.), optionally filter out sensitive information, and outport into a new PIM system.

What types and formats of data will the project generate/collect?

It depends on the format that the 3rd party systems provide their data. In most cases this are textual type with formats such as JavaScript Object Notation (JSON) and Comma Separated Values (CSV).

Will you re-use any existing data and how?

No.

What is the origin of the data?

3rd party Personal Information Management (PIM) systems (e.g. Facebook, Mobile Phone, Email etc.)

What is the expected size of the data?

It tends to be hundreds of megabytes per user.

To whom might it be useful ('data utility')?

This tool will benefit the users wishing to migrate their data to other platforms.

1.7. Fastweb

What is the purpose of the data collection/generation and its relation to the objectives of the project?

Fastweb will collect or process data in two ways:

1) Process data uploaded to PIMCity modules developed by other parties in the Consortium, because Fastweb will host some PIMCity modules on its cloud computing infrastructure. This has the purpose of testing the modules' functionality and enable interaction between modules and, potentially, third parties.

Collect aggregated network data like the number of users or traffic share of select websites or services. To be clarified.

What types and formats of data will the project generate/collect?

The type and format of data that will be processed in PIMCity modules running on Fastweb's cloud computing infrastructure will be determined later, together with the Consortium parties developing said modules.

Fastweb will collect aggregate network traffic data (e.g. traffic share of a particular website or service, bandwidth usage over time, number of users over time). The aggregation occurs in the dimension of users, i.e.: we will not collect or share any individual user identifier, but only user counts or session counts or bandwidth per website/service, possibly over time.

Will you re-use any existing data and how?

Fastweb might re-use some of the aggregated network traffic data it has been collecting for network capacity planning purposes, if useful for processing in PIMCity modules.

What is the origin of the data?

The data processed in PIMCity modules hosted in Fastweb's cloud computing infrastructure will come from the sources of said modules, developed and controlled by other Consortium parties. The data collected directly by Fastweb will come from network sensors installed in Fastweb's customer network.

What is the expected size of the data?

The size of the data will be determined later.

To whom might it be useful ('data utility')?

The data will be useful to the developers of PIMCity modules to enrich the data in their modules.

1.8. LSTech ESPANA

What is the purpose of the data collection/generation and its relation to the objectives of the project?

LSTech will provide applications/ mechanisms for data aggregation and anonymization and will define metadata schema for allowing the importing of personal data to the PIMCITY platform. LSTech does not need to collect or to have access to real personal data during the project. Test/ synthetic/ fake data can will be used to test the applications.

LSTech will also build user interfaces in the form of dashboards to allow the visualizations of the usage of personal data. If needed, LSTech will have access to the data (that will be collected and stored by other partners) in order to verify and test the efficiency and validity of these interfaces.

What types and formats of data will the project generate/collect?

LSTECH will use data collected and used by IMDEA since both are participating in the same task. The types and formats of data that will be collected/ used by IMDEA are mentioned in the respective section (1.2.4).

Data for the project more likely will include structured text.

Will you re-use any existing data and how?

LSTECH will use data collected and used by IMDEA since both are participating in the same task.

All the tests will be done by reusing public data, by downloading copies of the target datasets to local development environments and using them in the testbeds.

What is the origin of the data?

LSTECH will use data collected and used by IMDEA since both are participating in the same task.

Public data available in search engines (e.g. Google datasets) or provided by public entities (e.g. https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page). Such datasets are made public without including identifiable personal data for developers and data scientist to test their algorithms.

What is the expected size of the data?

It will strongly depend on the use case, ranging from MB to tens of GB (e.g. mobility data).

To whom might it be useful ('data utility')?

For the task T4.4, mobility data will be used by LSTECH and IMDEA.

Mobility data is useful for city planners or transportation enterprises at the time of planning resources or operations in their enterprises. CDR from mobile operators are being actively used in the billing process or for network planning purposes. Governments are using anonymized mobility information to track mobility in the city, for instance recently to measure the degree of mobility of the population during the Coronavirus outbreak.

1.9. KU Leuven – CiTiP

What is the purpose of the data collection/generation and its relation to the objectives of the project?

KU Leuven – CiTiP provides ethical and legal guidance to the other PIMCity partners. It does not collect personal data, does not collect or generate datasets, does not create algorithms, computer software and the like. KU Leuven – CiTiP collects and generates research data, however, which shall contribute to the fulfilment of the overall objectives of the Project.

What types and formats of data will the project generate/collect?

KU Leuven – CiTiP will generate research data in the form of deliverables which will be saved primarily in .docx and .pdf formats.

Will you re-use any existing data and how?

KU Leuven – CiTiP will re-use the generated research data for further research.

What is the origin of the data?

KU Leuven – CiTiP is generating research data relying on its own expertise in the subject matter.

What is the expected size of the data?

To be defined at the final stages of the project according to the documentation produced

To whom might it be useful ('data utility')?

Primarily for subjects researching privacy, data protection and intellectual property law.

1.10. Asociación de Usuarios de Internet

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The purpose of the data collected by AUI is to engage users to participate in the project by filling out surveys, participating in focusgroups and to participate as betatesters.

What types and formats of data will the project generate/collect?

Three types of data will be collected:

- Identifying data (name, email, phone) collected exclusively to contact participants who wish to give them voluntarily

- Sociodemographic data (age, gender, country, marital status, professional status, level of education, country, state, city, language) for the elaboration of studies and evaluation of the use of the different tools.

- Use and activity data (log files)

As for the formats these are collected in records and tables of mysql type databases in which the identification data are always stored encrypted.

Will you re-use any existing data and how?

No.

What is the origin of the data?

The data are provided by the users themselves when filling in the forms or registering on the website.

What is the expected size of the data?

The number of users will be less than 3000 and therefore the associated data can be less than 5 Megabytes.

To whom might it be useful ('data utility')?

For the set of partners that can through this data understand how different types of users perceive and iteract with the tools developed for them in the project.

1.11. Big Data Analytics

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The Trading Engine (TE) stores information about the Data Buyer, Offer, and transactions (Data Sellers that accepted to sell their Data), but not the Data itself. The primary purpose of storing it is transparency: The Seller can see what offers he has sold, to whom, and for what purpose. On the other side, the Buyer can see how many sellers have sold their data for the published offers.
The Taxonomy Generator (TG) receives as input raw personal data from different data sources and structures the data producing a schema definition for each one. The TG stores the schema definitions and exposes them through an API to let other processing systems consume them.

What types and formats of data will the project generate/collect?

The TE receives information coming from Data Sellers and Data Buyers as JSON, and it also responds to the transaction history in JSON format.

The raw input data format for the TG varies depending on the data source. For example, the historical geolocation data can be a KML or a JSON file. The schema generated for the data source is served as JSON.

Will you re-use any existing data and how?

The TE stores transactional Data for the duration of the action. During this period, data remains untouched, and both Data Sellers and Data Buyers can access it to see the history whenever they need to. The TG works similarly, receives raw input data, generates a schema, and serves it to the systems that need that information.

What is the origin of the data?

The TE stores data from the Offers created by Data Buyers and the transactions that happen when Data Sellers accept selling their data. The TG receives raw input data from the Data Portability and Control tool.

What is the expected size of the data?

The TE stores Offer and transactional data, and we expect to generate less than 10GB. The schema generated by the TG occupies less than 1GB.

To whom might it be useful ('data utility')?

The information generated by the TE is useful to Data Sellers and Data Buyers. The TG serves the schema generated to different systems that need that information.

1.12. CLIQZ

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The purpose of data collection by Cliqz is exclusive to satisfy the requiments of the project. Purely functional.

What types and formats of data will the project generate/collect?

On the scope of the consent manager (P-CM), the data will consist on tuples <identifier_of_user_data, consent_level> for each user. The consent_level is a structure common across all users. The identider_of_user_data is a pointer to arbitraty user's data stored on the P-DS.

On the scope of the Trading Engine (TE), the data will consist on contract, queries expresed by buyers. In the case of fullfillment of the contract, it will be stored for auditing purposed. Contracts contain no personal data but pointers to all relevant actors of it, namely: identider_of_user_data, identifier_of_seller, intefider_of_buyer and other metadata relevant for the contract such as date, experiration, price, etc.

Will you re-use any existing data and how?

The data gathered by Cliqz will not be re-used anywhere outside the scope of the project.

What is the origin of the data?

Data collected by Cliqz will exclusively come out of explicity input of the project's users.

What is the expected size of the data?

In the order of KB per user.

To whom might it be useful ('data utility')?

The data is only relevant to other components of the project.

2. FAIR DATA

a. Making data findable, including provisions for metadata

2.1.1. Politecnico di Torino

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

No. The data we plan to collect, process and produce is aggregated data.

What naming conventions do you follow?

To be defined.

Will search keywords be provided that optimize possibilities for re-use?

We will implement a searchable database to access and retrieve the privacy tags. The access to this data will be through open and standard API.

Do you provide clear version numbers?

Yes, we will identify the different versions of the privacy tag API with a version number.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

To be defined. No standard metadata exists for the privacy tags.
Similarly for Web archives, no standard metadata exists.

2.1.2. NEC Laboratories Europe

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

No.

What naming conventions do you follow?

It will be stored in a mysql database.

Will search keywords be provided that optimize possibilities for re-use?

No, we will not ask the user for consent to reuse.

Do you provide clear version numbers?

A single version will be generated.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

No.

2.1.3. Ermes Cyber Security

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

No. They are not.

What naming conventions do you follow?

It has not been defined yet.

Will search keywords be provided that optimize possibilities for re-use?

Privacy Metrics will be stored in a searchable database and made available through APIs.

Do you provide clear version numbers?

Yes, both Privacy Metrics and APIs will be versioned.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Metadata for Privacy Metrics have not been defined yet.

2.1.4. IMDEA Networks

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

The objective of the modules produced by IMDEA is to be used internally by other modules of the project. For instance, the relative values calculated for the different datasets contributing to a piece of traded data might be used by the trading engine in order to share the reward with data providers.
Regarding the watermarking, it just includes some tracking information in the data, but it does not alter the original metadata and/or its discoverability properties.

What naming conventions do you follow?

Any internal naming conventions defined in the project. No special naming conventions defined beyond those published in related papers to be released.

Will search keywords be provided that optimize possibilities for re-use?

N/A.

Do you provide clear version numbers?

N/A.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Metadata will be generated intended to be used internally within the project, as needed by other modules to reuse the results of the modules in charge of IMDEA.

2.1.5. Universidad Carlos III de Madrid

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

No. The data we plan to collect and process is aggregated data.

What naming conventions do you follow?

To be defined.

Will search keywords be provided that optimize possibilities for re-use?

The processed data will be offer through an API to third parties which help them to make the search for the value of the specific audience they are interested in.

Do you provide clear version numbers?

Yes, we will identify the different versions of the software with a version number.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

During the processing phase we can create metadata that allows to identify for instance ranges of value (min, max, median value), dynamics of the value evolution, etc.

2.1.6. Telefónica Investigación y Desarrollo

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

No.

What naming conventions do you follow?

To be defined.

Will search keywords be provided that optimize possibilities for re-use?

To be defined.

Do you provide clear version numbers?

No.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Since the aim is just to create a data portability tool for the user's benefit only, there will be no metadata created.

2.1.7. Fastweb

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

To be defined.

What naming conventions do you follow?

To be defined.

Will search keywords be provided that optimize possibilities for re-use?

To be defined.

Do you provide clear version numbers?

To be defined.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

To be defined.

2.1.8. LSTech ESPANA

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

LSTECH will provide mechanisms for analysing the personal data that will be collected/ stored for the project. Metadata are also available for these data. All these stored in the project platform “space”, not available for external entities, not discoverable by any means.

What naming conventions do you follow?

No special naming conventions. LSTECH will use the data as they will be defined/ stored in the project. The naming will be related to the context in order to facilitate better understanding and reusability of the data.

Will search keywords be provided that optimize possibilities for re-use?

N/A.

Do you provide clear version numbers?

In the services, applications, data, yes, where required, but not necessarily.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Metadata will be generated intended to be used internally within the project, as needed by other modules to use the results of the modules in charge of LSTECH.
Metadata standards do exist for some of the data that will be collected/ used in the project, like personal contact information, and they will be used for the project. We do not know if we will need to create more in this stage.

2.1.9. KU Leuven – CiTiP

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

Deliverables of KU Leuven – CiTiP include keywords and abbreviations that help navigating the research data in .docx and .pdf files.

What naming conventions do you follow?

N/A.

Will search keywords be provided that optimize possibilities for re-use?

Yes.

Do you provide clear version numbers?

Yes.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

KU Leuven – CiTiP uses keywords and abbreviations that help navigating the research data in .docx and .pdf files.

2.1.10. Asociación de Usuarios de Internet

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

No, the identification data is stored in encrypted form and the rest is stored disaggregated in different spaces in the database.

What naming conventions do you follow?

The data is stored in a MySQL database and the only limitation to the names is the implicit use of this tool.

Will search keywords be provided that optimize possibilities for re-use?

No.

Do you provide clear version numbers?

Yes.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

We do not use metadata.

2.1.11. Big Data Analytics

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

On the TE, a Data Seller can request his and only his transaction history. When the Data Seller accepts selling his Data to the Data Buyer, the TE registers this action using an identifier provided by the Seller that cannot be correlated with any personal data from him.
On the TG, all the generated schemas have a unique identifier with a naming convetion to be defined.

What naming conventions do you follow?

To be defined.

Will search keywords be provided that optimize possibilities for re-use?

On the TE, the search keyword would be the Data Sellers identifiers, and the Data Buyers also use their unique identifiers to fetch and see the progress of each of the Offers they published.
On the TG, the search keyword would be the schema identifier.

Do you provide clear version numbers?

On the TE, the system itself may evolve, and for that reason, it is versioned. The data generated and stored is immutable, and so there is no reason to have it versioned.
On the TG, the generated schemas may change as the raw input data could also change, so it is versioned.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Both systems do not generate metadata.

1.13. CLIQZ

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

No, the data which always come out of user's input should not be discoverable outside the scope of the project or needs to be standarized as it has no value outside the project itself.

What naming conventions do you follow?

To be defined with partners of the project.

Will search keywords be provided that optimize possibilities for re-use?

There is no use-case of discovery or search on the data collected by Cliqz of raw data. At the aggregate level data can be accessed through APIs, which are not public but restricted to users and partners of the project.

Do you provide clear version numbers?

The API will have a versioning mechanisms to ensure interoperability.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

The only metadata created byt the project that could be open would be the taxonomy, which contains no user-data in any shape of form. Only on that case, we should follow the standard of the base taxonomy of choice. Most likely, still to be decided, it will be the taxonomy of interests define by IAB.

b. Making data openly accessible

2.2.1. Politecnico di Torino

Which data produced and/or used in the project will be made openly available as the default?

This is still to be decided. Part or the whole data produced by POLITO in the project might be made openly public for its utilization. In particular, privacy tags will be made public accessible and searchable, while web crawling collections will be likely made available only upon specific requests due to the size of the data.
Since the produced rata will be aggregated data, and would not have any implication in individuals' data privacy.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

This is still to be decided. Part or the whole data produced by POLITO might be considered private and only shared under restrictions. This will depend on the commercial value of the produced data.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

The privacy tags will accessible through an API. The API can be queried.
Considering web archives, we could make (part of) it available for download on POLITO internal servers.

What methods or software tools are needed to access the data?

API will be based on web technologies.

Is documentation about the software needed to access the data included?

Yes, all documentation to use the API will be released by the project.

Is it possible to include the relevant software (e.g. in open source code)?

Yes, the PDK will include documented examples on how to access the API.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

To be agreed. In principle, we will make it available in the repository specified by the project.

Have you explored appropriate arrangements with the identified repository?

To be defined and agreed within the project. The consortium will grant the setup and maintenance of an internal repository.

If there are restrictions on use, how will access be provided?

No.

Is there a need for a data access committee?

No.

Are there well described conditions for access (i.e. a machine readable license)?

No.

How will the identity of the person accessing the data be ascertained?

To be defined once the repository of the project is set up. In principle, the usual ones based on accounts/userids. Part of the data may be open accessible.

2.2.2. NEC Laboratories Europe

Which data produced and/or used in the project will be made openly available as the default?

No data.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

We will have personal data, so, we will not share it.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

To be clarified.

What methods or software tools are needed to access the data?

To be clarified.

Is documentation about the software needed to access the data included?

To be clarified.

Is it possible to include the relevant software (e.g. in open source code)?

To be clarified.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

To be clarified.

Have you explored appropriate arrangements with the identified repository?

To be clarified.

If there are restrictions on use, how will access be provided?

To be clarified.

Is there a need for a data access committee?

To be clarified.

Are there well described conditions for access (i.e. a machine readable license)?

To be clarified.

How will the identity of the person accessing the data be ascertained?

To be clarified.

2.2.3. Ermes Cyber Security

Which data produced and/or used in the project will be made openly available as the default?

Data summarizing websites collected using web scrapers will not made available by ECS, while part or the Privacy Metrics produced by ECS in the project might be made public for its utilization using specific APIs. Sharing these data does not have any privacy implication.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Data collected by ECS using web scrapers represent a competitive advantage for ECS, and as such these data will not be shared or made publicly availale. Privacy Metrics generated by ECS might be made available under restrictions as ECS plans to exploit them commercially.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

Privacy Metrics will be made accessible through the use of specific APIs.

What methods or software tools are needed to access the data?

Whatever software tool capable of interacting with web APIs will be fine.

Is documentation about the software needed to access the data included?

Yes, all documentation to use the APIs will be released by the project.

Is it possible to include the relevant software (e.g. in open source code)?

Yes, the PDK will include documented examples on how to access the APIs.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

This has not been defined yet.

Have you explored appropriate arrangements with the identified repository?

This has not been defined yet.

If there are restrictions on use, how will access be provided?

Privacy Metrics made available by ECS in the context of the project will have not access restrictions.

Is there a need for a data access committee?

No.

Are there well described conditions for access (i.e. a machine readable license)?

No.

How will the identity of the person accessing the data be ascertained?

We will use accounts defined by the repository hosting platform.

2.2.4. IMDEA Networks

Which data produced and/or used in the project will be made openly available as the default?

As data from IMDEA Networks' modules is intended to be reused internally in the project and/or solution, it will not be openly available. The software will be part of the PDK, and thus will be made public as part of the deliverables of the project.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

N/A.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

No data will be made accessible. In case any demo is required we will link to the open public sources were data for the demo or example is downloadable.

What methods or software tools are needed to access the data?

N/A.

Is documentation about the software needed to access the data included?

There will, consequently, no specific tool to access any data, as no data will be shared. However, as part of PIMCITY deliverables, we will make available open source code and its related documentation for third parties to reuse. Part of this Open Source code will be libraries from IMDEA.

Is it possible to include the relevant software (e.g. in open source code)?

As part of PIMCITY deliverables, we will make available open source code and its related documentation for third parties to reuse. Part of this Open Source code will be libraries from IMDEA.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

The project website and potentially some websites like Github (to be explored) for open source shared code.

Have you explored appropriate arrangements with the identified repository?

We assume that data from the project will come from the project repository. It is internal from the project, hosted by Fastweb.

If there are restrictions on use, how will access be provided?

Not defined.

Is there a need for a data access committee?

Not defined.

Are there well described conditions for access (i.e. a machine readable license)?

Not defined.

How will the identity of the person accessing the data be ascertained?

Not defined.

2.2.5. Universidad Carlos III de Madrid

Which data produced and/or used in the project will be made openly available as the default?

This is still to be decided. Part or the whole data produced by UC3M in the project might be made openly public for its utilization.
Since the produced rata will be based on aggregated data, this does not have any implication in individuals' data privacy.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

This is still to be decided. Part or the whole data produced by UC3M might be considered private and only shared under restrictions. This will depend on the commercial value of the produced data.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

In principle data will be accessible through an API, although internally in the consortium alternative sharing methods can be discussed.

What methods or software tools are needed to access the data?

As indicated above, the data will accessible through an API. The API can be queried.

Is documentation about the software needed to access the data included?

Yes, documentation will be made available once the software is developed.

Is it possible to include the relevant software (e.g. in open source code)?

The software will be integrated within the PIMCITY infrastructure wherever it is determined in the design phase.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

In principle, we will make it available in the repository specified by the project.

Have you explored appropriate arrangements with the identified repository?

We are part of the consortium, so the arrangement is granted by the existence of the project.

If there are restrictions on use, how will access be provided?

No restrictions.
Is there a need for a data access committee?

No.

Are there well described conditions for access (i.e. a machine readable license)?

Once the repository from the project is set up, such conditions should be defined by third parties access.

How will the identity of the person accessing the data be ascertained?

To be defined once the repository of the project is set up. In principle, the usual ones based on accounts/userids.

2.2.6. Telefónica Investigación y Desarrollo

Which data produced and/or used in the project will be made openly available as the default?

No data will be made openly available by default.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Our tasks do not involve data sharing.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

No data will be made accessible since there will be no data sharing.

What methods or software tools are needed to access the data?

Data will be accessed by authorized parties only using the appropriate API that the 3rd party PIM provides. In cases that there is no API available, manual process might be considered on a case by case scenario.

Is documentation about the software needed to access the data included?

No data access will be provided, and thus no documentation is required.

Is it possible to include the relevant software (e.g. in open source code)?

As part of PIMCITY deliverables, we will make available open source code and its related documentation for third parties to reuse if authorized.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

As part of PIMCITY deliverables, we will make available open source code and its related documentation for third parties to reuse.

Have you explored appropriate arrangements with the identified repository?

We are part of the consortium, so the arrangement is granted by the existence of the project.

If there are restrictions on use, how will access be provided?

No restrictions.

Is there a need for a data access committee?

No.

Are there well described conditions for access (i.e. a machine readable license)?

To be defined by the Consortium.

How will the identity of the person accessing the data be ascertained?

No data will be committed to the repository.

2.2.7. Fastweb

Which data produced and/or used in the project will be made openly available as the default?

To be defined by the Consortium.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

To be defined. Personal traffic data from Fastweb customers cannot be shared, in compliance with privacy regulations, and we do not plan to use it in the project.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

To be defined by the Consortium.

What methods or software tools are needed to access the data?

To be defined by the Consortium parties developing PIMCity modules.

Is documentation about the software needed to access the data included?

To be defined by the Consortium parties developing PIMCity modules.

Is it possible to include the relevant software (e.g. in open source code)?

To be defined by the Consortium parties developing PIMCity modules.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

The data hosted in Fastweb's cloud computing infrastructure will be deposited in Fastweb data centres located in Italy.

Have you explored appropriate arrangements with the identified repository?

N/A.

If there are restrictions on use, how will access be provided?

To be defined by the Consortium parties developing PIMCity modules.

Is there a need for a data access committee?

To be clarified.

Are there well described conditions for access (i.e. a machine readable license)?

To be defined by the Consortium parties developing PIMCity modules.

How will the identity of the person accessing the data be ascertained?

To be defined by the Consortium parties developing PIMCity modules.

2.2.8. LSTech ESPANA

Which data produced and/or used in the project will be made openly available as the default?

N/A.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

N/A.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

No data will be made accessible.

What methods or software tools are needed to access the data?

N/A.

Is documentation about the software needed to access the data included?

N/A.

Is it possible to include the relevant software (e.g. in open source code)?

N/A.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

The project website and potentially some websites like Github (to be explored) for open source shared code.

Have you explored appropriate arrangements with the identified repository?

We assume that data from the project will come from the project repository. It is internal from the project, hosted by Fastweb.

If there are restrictions on use, how will access be provided?

Not defined.

Is there a need for a data access committee?

Not defined.

Are there well described conditions for access (i.e. a machine readable license)?

Not defined.

How will the identity of the person accessing the data be ascertained?

Not defined.

2.2.9. KU Leuven – CiTiP

Which data produced and/or used in the project will be made openly available as the default?

Research data, i.e. KU Leuven – CiTiP deliverables such as legal requirements.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

KU Leuven – CiTiP will not work with datasets.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

Primarily via project website.

What methods or software tools are needed to access the data?

N/A.

Is documentation about the software needed to access the data included?

No.

Is it possible to include the relevant software (e.g. in open source code)?

N/A.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

N/A.

Have you explored appropriate arrangements with the identified repository?

N/A.

If there are restrictions on use, how will access be provided?

N/A.

Is there a need for a data access committee?

No.

Are there well described conditions for access (i.e. a machine readable license)?

N/A.

How will the identity of the person accessing the data be ascertained?

N/A.

2.2.10. Asociación de Usuarios de Internet

Which data produced and/or used in the project will be made openly available as the default?

The data of the users participating in the project are not made public.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Contact details of users participating in the project are not shared. They are only used by UAI to contact the user if required at any stage of the project and to keep him/her informed of the project's progress.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

No.

What methods or software tools are needed to access the data?

Access to the database and knowledge of encryption keys.

Is documentation about the software needed to access the data included?

No.

Is it possible to include the relevant software (e.g. in open source code)?

No.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

The data is stored in the AUI's servers which are located in a datacenter operated by the hosting provider ARSYS.

Have you explored appropriate arrangements with the identified repository?

No.

If there are restrictions on use, how will access be provided?

No.

Is there a need for a data access committee?

No.

Are there well described conditions for access (i.e. a machine readable license)?

Yes.

How will the identity of the person accessing the data be ascertained?

Only people with system administrator status can access the databases and if they do, the identification data is always stored in encrypted form.

2.2.11. Big Data Analytics

Which data produced and/or used in the project will be made openly available as the default?

No data produced is shared openly.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Data produced is used inside the project.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

The information will be accessible through an API by other PIM systems.

What methods or software tools are needed to access the data?

The information can be accessed using a standard HTTP client library.

Is documentation about the software needed to access the data included?

Yes, both projects developed by Grandata include the documentation for APIs.

Is it possible to include the relevant software (e.g. in open source code)?

Yes, the documentation includes examples of how to query the information and these examples can be open-sourced. Whether both systems, the TE and TG, are open source is yet to be defined.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

To be defined. The repository includes code and documentation for sure, but no data will be stored there.

Have you explored appropriate arrangements with the identified repository?

To be discussed.

If there are restrictions on use, how will access be provided?

To be discussed.

Is there a need for a data access committee?

To be defined.

Are there well described conditions for access (i.e. a machine readable license)?

To be defined.

How will the identity of the person accessing the data be ascertained?

It depends on the service used by the project to host the repository. Again, no data will commited to the repository.

1.14. CLIQZ

Which data produced and/or used in the project will be made openly available as the default?

The only data to be made public is the taxonomy of intents and types of data (the identifier_user_data). Still to be decided, but most likely we will extend the taxonomy of user's interests of IAB.

If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

No data collected by Cliqz needs to be made public as it only applies to the scope of the project. The only exception would be the aforementioned taxonomy (not build solely by Cliqz but with all partners of the project).

Users might decide to export their consent preferences to other systems, which they should be able to do through the API. We will guarante data-portability of any data that the user did input to our project (via an API).

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

The taxonomy should exists on a repository, accessible online.

Data portability for users should be based on an API, which only users can access or provide access to.

What methods or software tools are needed to access the data?

To be able to access a REST API.

Is documentation about the software needed to access the data included?

All APIs will be fully documented.

Is it possible to include the relevant software (e.g. in open source code)?

As part of the project requirements, all relevant software will be made available open-source.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

To be defined, but most likely a version control like GIT will be used.

Have you explored appropriate arrangements with the identified repository?

To be defined.

If there are restrictions on use, how will access be provided?

None that we are aware of. For Cliqz related code and data, there is no need for one.

Is there a need for a data access committee?

None that we are aware of. For Cliqz related code and data, there is no need for one.

Are there well described conditions for access (i.e. a machine readable license)?

Not applicable.

How will the identity of the person accessing the data be ascertained?

Only projects of users that have sign-in to the project can access data.
The expection should be the taxonomy, which can/should publicly accessible without restrictions (with an open-source license).

c. Making data interoperable

2.3.1. Politecnico di Torino

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

This needs to be defined.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

This needs to be defined.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

This needs to be defined.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Yes, if needed and possible.

2.3.2. NEC Laboratories Europe

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

This needs to be defined.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

This needs to be defined.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

This needs to be defined.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

To be clarified.

2.3.3. Ermes Cyber Security

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

This needs to be defined.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

This needs to be defined.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

[This needs to be defined.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Yes, if needed.

2.3.4. IMDEA Networks

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

Yes. We will be using Python and open software and libraries, as well as data formats to produce the results of the investigation. Even we plan to share such data and developments with papers produced throughout the project.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Not defined.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Not defined.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

To be clarified.

2.3.5. Universidad Carlos III de Madrid

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

We are not aware of other datasets of the nature of the one to be produced by UC3M. So interoperability will need to be explored once datasets of similar nature are identified. If this happens during the execution of the project we will analyse this aspect.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

N/A.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

N/A.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

There will be a clear documentation associated to the developed tool that will allow anyone with the appropriate permission to access the produced data.

2.3.6. Telefónica Investigación y Desarrollo

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

There will be no data sharing available.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

There will be no data sharing available.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

There will be no data sharing available.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

N/A.

2.3.7. Fastweb

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

To be defined by the Consortium parties developing PIMCity modules.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

To be defined by the Consortium parties developing PIMCity modules.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

To be defined by the Consortium parties developing PIMCity modules.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

To be clarified.

2.3.8. LSTech ESPANA

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

It is a decision and action of the whole consortium. At the moment LSTECH does not have specific plans to produce data to be exchanged.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Not defined.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Not defined.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

N/A.

2.3.9. KU Leuven – CiTiP

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

Yes. KU Leuven – CiTiP deliverables which include research data are easily accessible with word processing programmes.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

N/A.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

N/A.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

N/A.

2.3.10. Asociación de Usuarios de Internet

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

To be clarified.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

To be clarified.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

To be clarified.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

2.3.11. Big Data Analytics

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

For the TG, schema definitions generated are served to different systems such as the Personal Data Safe or the Personal Consent Manager.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Yet to be aligned within the partners of the project.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Yes, but it needs to be defined.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Grandata will align to more commonly used ontologies, and document what is not aligned.

2.3.12. CLIQZ

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

Aggregated data might be useful for researchers. Both the user contents, to get statistics of the willingness of users to share certin type of data. Aggregates on the contracts of the trading engine would also be useful for researchers to assess market value.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

We are not aware of any standard vocabulary for the shareable data we could provide. Note, however, that the taxonomy, which will be extend is already a de-facto industry standard for advertising.

If there was a more general standard, we would consider using it.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

That would be the taxonomy, which will be an extension on top of the de-factor standard provided by IAB.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

We will put emphasis on using standarization where possible, and for the case of project specific knowledge, it will be documented so that it can either become a standard or it becomes easily re-usable.

d. Increase data re-use (through clarifying licences)

2.4.1. Politecnico di Torino

How will the data be licensed to permit the widest re-use possible?

This will be defined later in the project once we have more information and a better judgment to make the decision. For instance, if the data has commercial value, it will be probably licensed through commercial licenses. Another option is to license the data in a free manner. Finally a combination of the two options is also foreseen.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Possibly yes for web archives.
To be defined for the privacy tags, based on the exploitation plan of the project and partners.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Yes, the data is usable by third parties also after the end of the project. This use will be ruled through the specific licenses once defined.

How long is it intended that the data remains re-usable?

For at least 5 years.

Are data quality assurance processes described?

To be defined.

2.4.2. NEC Laboratories Europe

How will the data be licensed to permit the widest re-use possible?

To be clarified.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

To be clarified.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

To be clarified.

How long is it intended that the data remains re-usable?

To be clarified.

Are data quality assurance processes described?

To be clarified.

2.4.3. Ermes Cyber Security

How will the data be licensed to permit the widest re-use possible?

This will be defined later in the project.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

It is too early to define deadlines at this state of the project.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Since ECS is a commercial company, Privacy Metrics generated by ECS within the project might be usable by third parties under restrictions (technical, commercial and legal) which will be defined later in the project.

How long is it intended that the data remains re-usable?

To be defined.

Are data quality assurance processes described?

To be defined.

2.4.4. IMDEA Networks

How will the data be licensed to permit the widest re-use possible?

No data will be provided, except the data and algorithms useful to reproduce the research conducted by IMDEA.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

N/A.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

N/A.

How long is it intended that the data remains re-usable?

N/A.

Are data quality assurance processes described?

N/A.

2.4.5. Universidad Carlos III de Madrid

How will the data be licensed to permit the widest re-use possible?

This will be defined later in the project once we have more information and a better judgment to make the decision. For instance, if the data has commercial value, it will be probably licensed through commercial licenses. Another option is to license the data in a free manner. Finally a combination of the two options is also foreseen.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

The data will only be made available for re-use once the internal goals of UC3M and PIMCITY that requires the use of data in exclusivity are achieved.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Yes, the data is usable by third parties further than the end of the project. This use will be ruled through the specific license that is decided to apply to the data.

How long is it intended that the data remains re-usable?

This is hard to predict. If the data produced shows a clear commercial value, it will be remain re-usable for an undefined period of time until it loses its commercial value. Similar principles apply for research use of the data.

Are data quality assurance processes described?

We will conduct data quality assurance tests and will be properly reported when it is due.

2.4.6. Telefónica Investigación y Desarrollo

How will the data be licensed to permit the widest re-use possible?

No data will be shared.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

No data will be shared.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

No data will be shared.

How long is it intended that the data remains re-usable?

No data will be shared.

Are data quality assurance processes described?

No data will be shared.

2.4.7. Fastweb

How will the data be licensed to permit the widest re-use possible?

To be defined by the Consortium parties developing PIMCity modules.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

To be defined by the Consortium parties developing PIMCity modules.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

To be defined by the Consortium parties developing PIMCity modules.

How long is it intended that the data remains re-usable?

To be defined by the Consortium parties developing PIMCity modules.

Are data quality assurance processes described?

To be defined by the Consortium parties developing PIMCity modules.

2.4.8. LSTech ESPANA

How will the data be licensed to permit the widest re-use possible?

No data will be provided.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

N/A.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

N/A.

How long is it intended that the data remains re-usable?

N/A.

Are data quality assurance processes described?

N/A.

2.4.9. KU Leuven – CiTiP

How will the data be licensed to permit the widest re-use possible?

KU Leuven deliverables will not be licensed.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

As soon as possible; see general approach to data management as identified in the data management plan.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

See general approach to data management as identified in the data management plan.

How long is it intended that the data remains re-usable?

See general approach to data management as identified in the data management plan.

Are data quality assurance processes described?

See general approach to data management as identified in the data management plan.

2.4.10. Asociación de Usuarios de Internet

How will the data be licensed to permit the widest re-use possible?

N/A.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

N/A.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

No.

How long is it intended that the data remains re-usable?

N/A.

Are data quality assurance processes described?

N/A.

2.4.11. Big Data Analytics

How will the data be licensed to permit the widest re-use possible?

To be defined.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

To be defined.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

On the TE, all the information generated served the Data Sellers and Data Buyers for historical purposes. Data is not publicly available so it not available for re-use at the moment.
The TG generates schema definitions from the different data sources and it may be re-used by other projects but it needs to be defined.

How long is it intended that the data remains re-usable?

To be defined.

Are data quality assurance processes described?

To be defined.

2.4.12. CLIQZ

How will the data be licensed to permit the widest re-use possible?

For aggregated data, valuable only for research, we will open it through an open-source license to be hosted on a repository. To be discussed which one.

For non aggregated data, the data can only be shared by or with the permission of the users who input that data (through an API). Also covering for data-portability.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

To be defined. Data portability should be from day one.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Do no see the need for restriction on aggregated data provided by Cliqz. For data provided by Cliqz but whose origin is the user input the restrictions are imposed by the user itself.

How long is it intended that the data remains re-usable?

Aggregated, indefinitedly. User specific data whatever is the legal obligation of retention but not longer.

Are data quality assurance processes described?

To be defined.

3. ALLOCATION OF RESOURCES

What are the costs for making data FAIR in your project?

Costs are mostly represented by two main contributions:
1. Costs to manage and make data accessible for a Fair approach and
2. Costs for hosting data API in servers that hosts data and offer API to access it.

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Each partner is responsible for the costs incurred in making data FAIR, and for open access.

Who will be responsible for data management in your project?

Responsible partners are indicated in the table at the beginning of this document. In particular, specific roles are foreseen for technical coordinator (T8.2); dissemination manager (T6.1); innovation manager (T6.4); data manager (T7.1). Besides, PIMCity data protection officer team, subject to their competences, will advise the PIMCity project partners on data management.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

To be defined and agreed in the future.

4. DATA SECURITY

4.1. Politecnico di Torino

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

POLITO does not store or transfer sensitive data in the context of PIMCITY.
In general, all data collected and produced by POLITO will be stored in internally hosted datacenters, protected by standard mechanisms like firewalls. The access to the servers and services is based on personal authentication via credentials. Each access is logged.

Is the data safely stored in certified repositories for long term preservation and curation?

No – and likely not in the near future.

4.2. NEC Laboratories Europe

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

All data is tranferred over encrypted connections.

Is the data safely stored in certified repositories for long term preservation and curation?

The data will be stored at NEC premises. The computer where the data is stored is located in a room where only authorized personel can acces. The computer where the data is stored does not have direct connection to the Internet.

4.3. Ermes Cyber Security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

In general, ECS folows rather strict policy for data security. See DPIA for details. In general,
• all data collected and produced by ECS are stored and transferred using strong encryption;
• data are stored on both internal servers and private remote datacenters in multiple copies for redundancy;
• all accesses are protected by firewalls, logged and allowed to authorized personnel only.

Is the data safely stored in certified repositories for long term preservation and curation?

Yes, all data collected by ECS are safely duplicated in private remote datacenters.

4.4. IMDEA Networks

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

No sensitive data will be stored or transferred. We will only use open data available in the internet without any “personal” information.

Is the data safely stored in certified repositories for long term preservation and curation?

It will be stored in the computers used in the research process.

4.5. Universidad Carlos III de Madrid

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

The data will protected using start-of-the-art security practices. It will be stored in servers at UC3M which provide a full protected network with firewall, traffic monitoring, server access control (user, passwords, encryption options), servers located in rooms with physical access control, periodic data backups, etc.

If the data is moved to a different repository based on project’s decision, similar security provisions will be demanded.

Is the data safely stored in certified repositories for long term preservation and curation?

Not that we are aware of.

4.6. Telefónica Investigación y Desarrollo

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

All data will be stored encrypted using state-of-the-art encryption technology. Even if the stored data gets lost, the user can request another data importation from the original sources and restore the data.

Is the data safely stored in certified repositories for long term preservation and curation?

No.

4.7. Fastweb

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Fastweb’s cloud infrastructure has all the security, privacy and continuity provisions included in the international standards for information security (ISO 27001), business continuity (ISO 22301), IT service management (ISO 20000), cloud services security (ISO 27017), cloud service privacy (ISO 27018), and security incident management (ISO 27035). Fastweb holds certifications for the aforementioned standards, issued by accredited third parties.

Is the data safely stored in certified repositories for long term preservation and curation?

To be clarified.

4.8. LSTech ESPANA

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

LSTECH is not storing data for the project at its own premises. No sensitive data will be stored or transferred.

To be clarified.

Is the data safely stored in certified repositories for long term preservation and curation?

It will be stored in the computers used in the research process.

To be clarified.

4.9. KU Leuven – CiTiP

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

To be clarified.

Is the data safely stored in certified repositories for long term preservation and curation?

To be clarified.

a. Asociación de Usuarios de Internet

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

N/A.

Is the data safely stored in certified repositories for long term preservation and curation?

N/A.

4.10. Big Data Analytics

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Information on the supporting assets is backed up once every day, and the amount of backups held is limited to 1 month. Assets supporting personal data are not publicly available and are part of a Virtual Private Network accessible only from the application servers. The channel for interacting with the applications is securely encrypted.

Is the data safely stored in certified repositories for long term preservation and curation?

No.

4.11. CLIQZ

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Storage of data is on servers behind a VPN, access to it is only though a secure API with HTTPS endpoints.

For publicly shareble data, such as high-level aggregates or the taxonomy, only integrity is needed, which can be done by signing the fingerprint of the data with a project certificate. This, however, needs to be discussed at it should be agreed across all partners.

Is the data safely stored in certified repositories for long term preservation and curation?

For operational data, the security provided by the backend servers and storage (behind a VPN) should suffice.

For public data, we would use a certified repository for long-term access (beyond the lifetime of the project).

5. ETHICAL ASPECTS

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Please see deliverables D7.1, D7.2, D9.1, D9.2.

5.1. Politecnico di Torino

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

We do not foresee to collect personal data. In case, yes – we will consider explicitly the collection of informed consent for data sharing and long term preservation.

5.2. NEC Laboratories Europe

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

No.

5.3. Ermes Cyber Security

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

We will not collect personal data in the context of the project.

5.4. IMDEA Networks

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

N/A.

5.5. Universidad Carlos III de Madrid

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

We do not plan to collect any personal data.

5.6. Telefónica Investigación y Desarrollo

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

No.

5.7. Fastweb

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

As of April 2020, Fastweb does not plan to collect personal data.

5.8. LSTech ESPANA

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

N/A.

5.9. KU Leuven – CiTiP

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

N/A.

5.10. Asociación de Usuarios de Internet

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

Consent is requested on all forms where users are required to provide personal data and are informed of the use that will be made of it and their rights in accordance with the GDPR.

5.11. Big Data Analytics

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

No.

5.12. CLIQZ

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

There will be no sharing of any data that would be subjected to informed consent. Any transfer of data that would require such action will be initiated by the user itself (e.g. data-portability).

Taxonomy is built without involvement of user's data.

6. OTHER

6.1. Politecnico di Torino

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

We follow the best practice that are commonly used in the research community.

6.2. NEC Laboratories Europe

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

We follow the IMSS policies of NEC.

6.3. Ermes Cyber Security

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

ECS is completing the process to get ISO 9001 and 27001 certifications.

6.4. IMDEA Networks

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

To be clarified.

6.5. Universidad Carlos III de Madrid

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

We follow best practices from the research community for data collection, processing and management.

6.6. Telefónica Investigación y Desarrollo

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

Any use/access/processing of data follows internal departmental policies of TID.

6.7. Fastweb

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

We follow industry best practices and comply with Italian laws concerning the generation, handling and deletion of data.

6.8. LSTech ESPANA

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

N/A.

6.9. KU Leuven – CiTiP

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

To be clarified.

6.10. Asociación de Usuarios de Internet

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

To be clarified.

6.11. Big Data Analytics

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

No.

6.12. CLIQZ

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

We follow the best practice that are commonly used in the research community.