Tailored IoT & BigData Sandboxes and Testbeds for Smart, Autonomous and Personalized Services in the European Finance and Insurance Services Ecosystem

Infinitech

D1.3 – Data Management Plan

Lead Beneficiary	**GFT**
Due Date	2020-03-31
Delivered Date	2020-03-31
Revision Number	3.0
Dissemination Level	Public (PU)
Type	Report (R)
Document Status	Release
Review Status	Internally Reviewed and Quality Assurance Reviewed
Document Acceptance	WP Leader Accepted and Coordinator Accepted
EC Project Officer	Pierre-Paul Sondag

HORIZON 2020 - ICT-11-2018

This project has received funding from the European Union's horizon 2020 research and innovation programme under grant agreement no 856632

Contributing Partners

Partner Acronym	Role[1]	Name Surname[2]
GFT	Authors	Elisa Rossi, Maurizio Megliola, Ernesto Troiano

Revision History

Version	Date	Partner(s)	Description
0.1	2020-01-30	GFT	ToC Version
1.0	2020-03-20	GFT	First Version for Internal Review
2.0	2020-03-27	GFT	Version for Quality Assurance
3.0	2020-03-31	GFT	Version for Submission

[1] Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance

[2] Can be left void

Executive Summary

Deliverable D1.3 – Data Management Plan (DMP) contains the INFINITECH Project's data management life cycle for the data to be produced, collected, processed, stored and preserved within the project development and beyond. In this action, we envision the following different types of data:

- data provided by the use cases,
- publications (e.g., conference papers),
- public deliverables,
- confidential deliverables,
- open source software,
- closed source software,
- artefacts of research value produced by application executions (e.g., logs, playbooks),
- working documents of the project.

Following the EC template[3], this document presents how these different types of data will be collected, who the main beneficiaries are, and how they will be stored and managed within the project, and if the project will make them accessible, findable and re-usable.

The DMP includes for each category tools and procedures to manage data according to policy and regulation. In particular the DMP is based on:

- A data Taxonomy,
- A Data registry and risk management.

The deliverable describes the resources needed for the openness and data to finalize with security and ethical aspects that will be taken into consideration in the context of INFINITECH.

[3] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Table of Contents

List of Figures

List of Tables

Abbreviations

FAIR	Findable, Accessible, Interoperable and Re-usable
MiFID	Markets in Financial Instruments Directive
MiFIR	Markets in Financial Instruments and Amending Regulation
NDA	Non-Disclosure Agreement
NIS	Network and Information Systems
OES	Operators of Essential Services
PAN	Primary Account Number
PaaS	Platform as a Service
PCI DSS	Payment Card Industry Data Security Standard
PIA	Privacy Impact Assessment
PSD2	Payment Service Directive 2
PSP	Payment Service Provider
PSU	Payment Service User
P2PP	Peer-to-Peer Payment
RTS	Regulatory Technical Standard
QTSP	Qualified Trust Service Provider
SCA	Strong Customer Authentication
SME	Small and Medium-Sized Enterprises
SA	Supervisory Authority
SECaaS	Security-as-a- Service
TI	Threat Intelligence
3DS	Three-Domain Secure

1 Introduction

This deliverable focuses on the management of the data in the INFINITECH project.

The following kind of data will have to be handled:

- The first kind of data **are the datasets that will be provided by the project's pilots (in accordance with the project's DoA there will be fifteen (15) pilots in total)** and that will be used to validate the project. For each use case, the respective data has its specific requirements. Although anonymized, these data sets are consortium confidential. Moreover, they may be complemented by open data sets.
- The second kind of data are the **publications**, either scientific or technical, that are based on the findings of the project.
- The third and fourth kind of data are the project's **deliverables,** which are marked as 'Public', in which case they will be accessible from the Project Web site, or 'Confidential', in which case their content is accessible only within the consortium and the European Commission's Agency funding the project.
- The fifth kind of data are the **software artefacts,** they could be released as open software or **proprietary (**closed).
- **Artefacts of research value obtained from the INFINITECH infrastructure**. For instance, logs or playbooks generated by applications being executed over INFINITECH may be of research interest.
- Working documents, i.e. mainly working versions of project deliverables, stored in a shared repository based on G-Suite

As part of making research data findable, accessible, interoperable and re-usable (FAIR[4]), this deliverable includes information with regard to:

- the handling of research data during & after the end of the project,
- 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).

1.1 Objective of the Deliverable

The objective of the deliverable is to detail and regulate the data management of all the aforementioned categories of data to be produced, collected, processed, stored and preserved within the INFINITECH project development and beyond.

1.2 Insights from other Tasks and Deliverables

The present deliverable is directly related to other workpackages and tasks in particular with WP2 for Data Asset specification (task T2.5) and WP7 - Large-Scale Pilots of SHARP Financial and Insurance Services, where the fifteen (15) pilots of the project will define, design, develop and evaluate their use cases by using the INFINITECH solution.

1.3 Structure

The document follows the established H2020 template for a Data Management Plan (DMP).[5]

[4] https://www.go-fair.org/fair-principles/

[5] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

2 Data Summary

INFINITECH project aims at providing novel BigData/IoT solutions for seamless management and querying of all types of data (e.g., OLAP/OLTP[6], structured/unstructured/semi-structured, data streaming & data at rest), blockchain-based data sharing and real-time analytics in a broad range of business cases in the Financial Sector. To facilitate compliance to regulations both specific for the banking sector (e.g., PSD2[7], 4AMLD[8], MIFiD II[9]) and compliance with other EU regulations (e.g., GDPR[10]), regulatory tools incorporating various data governance capabilities (e.g. anonymization, eIDAS[11] integration) will be available for the pilots and demonstrators.

From this brief overview of the objectives of the project, there is a clear need for proper data management at all the levels of the Big Data Value Chain[12].

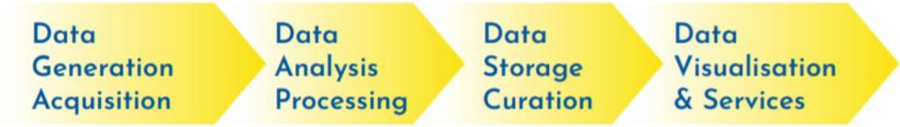

Figure 1 - Big Data Value Chain defined by BDVA[12]

The Consortium, to adhere at best to status of the art of Big Data Models, refers to the model proposed by the BDVA (Big Data Value Association).

The BDV Reference Model[12] has been developed by the BDVA and serves as common reference framework to locate Big Data technologies on the overall IT stack. It addresses the main concerns and aspects to be considered for Big Data Value systems.

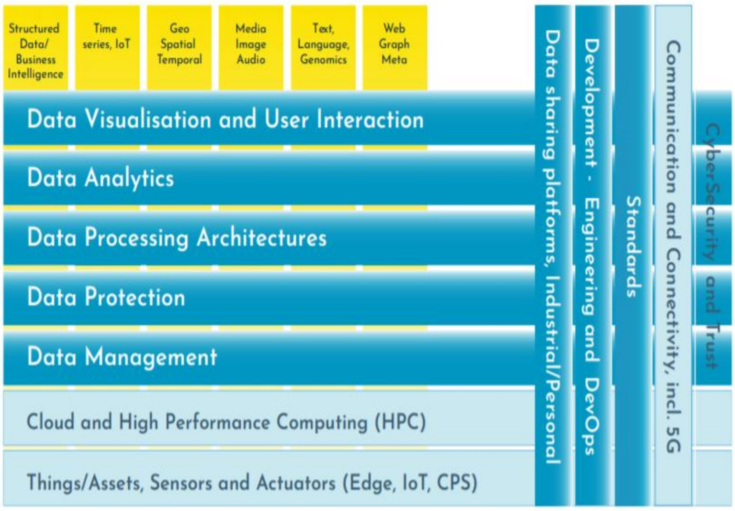
Structured Data/ Business Intelligence
Time series, IoT
Geo Spatial Temporal
Media Image Audio
Text, Language, Genomics
Web Graph Meta
Data Visualisation and User Interaction
Data Analytics
Data Processing Architectures
Data Protection
Data Management
Cloud and High Performance Computing (HPC)
Things/Assets, Sensors and Actuators (Edge, IoT, CPS)
Data sharing platforms, Industrial/Personal
Development - Engineering and DevOps
Standards
Communication and Connectivity, incl. 5G
Cyber Security and Trust

Figure 2 - Big Data value reference model[12]

[6] Respectively, Online Analytical Processing and Online transaction processing databases

[7] https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366_en/

[8] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32015L0849

[9] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:145:0001:0044:EN:PDF

[10] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

[11] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG

[12] http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf

The BDV Reference Model is structured into horizontal and vertical concerns:

- Horizontal concerns cover specific aspects along the data processing chain, starting with data collection and ingestion, and extending to data visualisation. It should be noted that the horizontal concerns do not imply a layered architecture.
- Vertical concerns address cross-cutting issues, which may affect all the horizontal concerns. In addition, vertical concerns may also involve non-technical aspects.

2.1 Data provided by the use cases

The data provided, used and produced by the pilots of the project will be handled in full compliance of the main legislation and directives and more specifically:

- The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data[13],
- The European Convention for the Protection of Human Rights and Fundamental Freedoms[14],
- Directive 95/46/EC[15] & Directive 2002/58/EC[16] of the European parliament regarding issues with privacy and protection of personal data and the free movement of such data,
- Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use[17],
- Council Directive 83/570/EEC of 26 October 1983 amending Directives 65/65/EEC, 75/318/EEC and 75/319/EEC on the approximation laid down by law, regulation or administrative action relating to proprietary medicinal products[18],
- Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the legal protection of biotechnological inventions[19],
- Directive on Privacy and Electronic Communications (2002/58/EC)[20],
- Directive on Protection of Privacy in the Telecommunication Sector (97/66/EC)[21],
- DIRECTIVE 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA[22].

In addition, INFINITECH will fully comply to specific (and in certain cases more strict) national legislation of the pilot countries involved in the project.

[13] https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108

[14] https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005

[15] https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A31995L0046

[16] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058

[17] https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/dir_2001_20/dir_2001_20_en.pdf

[18] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31983L0570

[19] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31998L0044

[20] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32002L0058

[21] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31997L0066

[22] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0680

The compliance with principles of data protection will be applied to all the steps of the Big Data Value Chain (Figure 1) as well as to all the layers defined in the Big Data value reference model (Figure 2).

2.1.1 Data Taxonomy

The following schema (Figure 3) shows the three typologies of data that will be acquired, used, analysed and produced by the fifteen (15) pilots of INFINITECH within the WP7 - Large-Scale Pilots of SHARP Financial and Insurance Services.

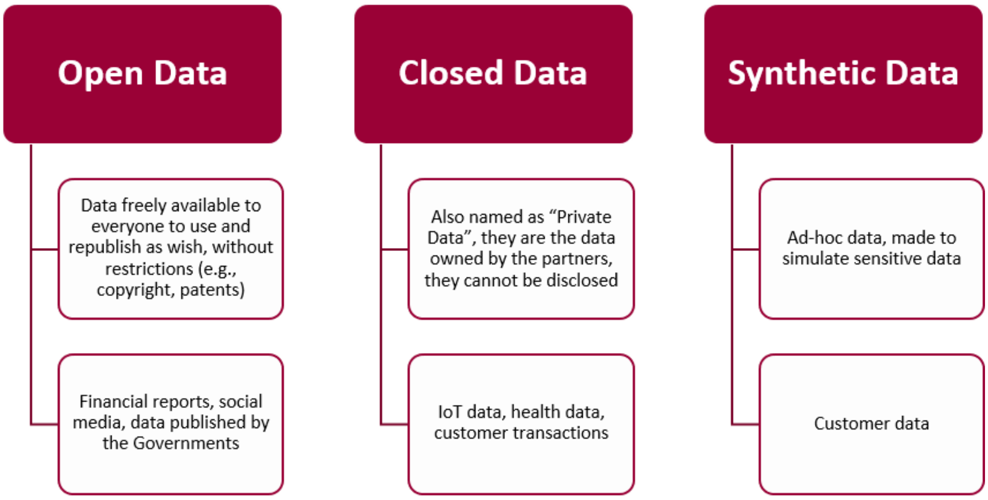
Open Data
Data freely available to everyone to use and republish as wish, without restrictions (e.g., copyright, patents)
Financial reports, social media, data published by the Governments
Closed Data
Also named as "Private Data", they are the data owned by the partners, they cannot be disclosed
IoT data, health data, customer transactions
Synthetic Data
Ad-hoc data, made to simulate sensitive data
Customer data

Figure 3 - Taxonomy of the data used by the INFINITECH pilots, definition and examples

For each of the three macro-categories identified in Figure 3, Open Data, Closed Data and Synthetic Data, the FAIR (Findable, Accessible, Interoperable, Reusable) principles are detailed in the next table (Table 1).

Table 1 - FAIR principles for the pilots Data Management in INFINITECH

	Open Data	**Closed Data**	**Synthetic Data**
Making data findable, including provisions for metadata	They are already available on the Internet, the pilots will specify time by time the source.	These data cannot be disclosed, eventually the results of their analysis will be disclosed as aggregated and anonymized data.	The subsets of data that will be made openly accessible will be well detailed through metadata.
Making data openly accessible	They are already accessible.	The pilots will evaluate case by case the possibility to make the data openly accessible.	The pilots will be encouraged to share in ad-hoc websites at least a subset of the synthetic data produced for the execution of the use cases.
Making data interoperable	They should be already in the most common formats (e.g., .csv, .xls).	Data have to be interoperable in order to ease their use	The interoperability of the data is guaranteed as it is needed also

		combined with other data (e.g., open data).	from the project perspective.
Increase data re-use (through clarifying licenses)	These data are not licensed.	The re-use of the reports/datasets produced by the analytics on the closed data will be evaluated case by case by the pilots.	The re-use of the reports/datasets produced by the analytics on the closed data will be evaluated case by case by the pilots.

Furthermore, even if the Closed Data will be anonymized before their usage within the pilots' scope, in the case of sensitive data the principle of data minimization will be applied.

2.1.2 Pilot Datasets

In WP2 T2.5 a first dataset for the pilot are assessed. This process includes gathering info on:

- Dataset Provider
- Dataset Name
- Dataset (short) description
- Owner
- License/Privacy
- Anonymized
- Capability of Synthetic Data Production
- Data Type
- Data format
- Estimated Storage needs
- Dataset Version
- Coverage (From-To Years)
- Data store
- Recommended API
- Data Volume
- Data Velocity
- Language
- Availability
- Metadata available (y/n)
- Data Searchable with keywords (y/n)
- Standard Vocabularies for interoperability (y/n)
- What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
- Data availability after INFINITECH (y/n)
- Link to Dataset Documentation
- Link to Sample
- Link to Complete Dataset (if possible)
- Responsible (Leader Contact Person) & e-mail
- Proxy (Tech Advisor) & e-mail

These information per Pilot are kept in a registry on the repository (EXCEL FILE) and the information protected in the same way in the case of other project documents (cfr. Section 2.6).

2.1.3 Publications

In the H2020 framework programme, publications (e.g., conference and journal peer review papers) are required to be openly accessible. All the partners have been made aware of this and are expected, if relevant, to pay the fees to publishers needed so that their publications are openly accessible.

2.2 Deliverables

Public Deliverables

All the deliverables which are marked with Public access have to and will be put online in the project website (https://www.infinitech-h2020.eu/).

Confidential deliverables

All the deliverables which are marked with Confidential will be reserved and accessible to the INFINITECH partners, the reviewers and EC (European Commission). The procedures are defined in the D1.1 – Project Reference Manual.

2.3 Software artefacts

Open Source

The Open Software artefacts that are produced in relation to the project will be of one of the following types:

- Open source code up streamed to big open source projects.
- Open source code that will be made available through open repositories (e.g. Github) allowing its utilization by other entities (e.g. stakeholders, researchers, etc) following partners' exploitation paths and plans.

Proprietary (Closed) Software

The IPRs of the closed software and applications are regulated by the Consortium Agreement and License Agreement.

2.4 Artefacts of research value produced by application executions

Various components of the INFINITECH infrastructure when running big data applications produce artefacts (e.g., logs, playbooks, etc) that may be of research interest. This is applicable for logs, playbooks, real-time information about the QoS of running applications. As of February 2020, it is premature to state to what extend such artefacts will be made available. In case we will have research quality artefacts, they will be packaged (e.g., zipped) and uploaded on software repositories of the project. This will permit open access and reuse of such artefacts.

The collected data content may be of interest to both the commercial sectors from which they were collected, as well as to a wider community of data scientists, or students of data science, to carry out machine learning research. Big Data infrastructure developers may be interested in the data as a means of performance testing against large volumes of data.

2.5 Working documents

The INFINITECH project maintains a document and file repository to provide a common archive where to store working documents related to the project, in particular the material to produce documentation and the Deliverables to be sent to the European Commission.

2.5.1 Security Policies

Disclaimer: the repository is NOT intended to store confidential information or IP materials belonging exclusively to organizations. The repository will store documents with different kind Information Classification: **Public, Confidential** or **Restricted** applying access control policies.

Therefore, a great deal of attention is dedicated to secure the infrastructure from unauthorized access to respect information classification, privacy, ethical and legal frameworks in particular the GDPR and the internal security policies of the Project Coordinator (GFT compliance) and other beneficiaries. In cases that the repository policies are found to be in conflict with the stakeholders' policies, these former frameworks will prevail.

2.5.2 Project Security Responsible / Data Protection Officer

GFT will appoint a **Project Security Responsible** (**PSR**) for security. The PSR is responsible for supplying access to users, deciding about access rights to specific assets, periodically reviewing the policies in place, getting reports about data breaches, etc.

A **Data Protection Officer** will be appointed by each Partner as this is also requested by the Ethical Requirements of the Grant Agreement. The DPO will assess the compliance with internal and legal security procedures.

2.5.3 Access Control

The access control policy is as follows:

1. Every partner will appoint one or more users responsible to access the repository
2. Every user will have assigned a unique account to access the repository
3. The credentials of the users will be supplied by **PSR**
4. Access is granted by use of a secure password and Strong Authentication (e.g. 2FA)
5. Repository folders are shared ONLY to authorized accounts within the project
6. Folders will be shared among Partners based on responsibility per WP, Task, Deliverable
7. Users will be responsible of keeping their credentials safe
8. Any data breach must be reported to the **Project Security Responsible** emailing at databreach@infinitech-h2020.eu
9. Project Security Responsible will take actions to mitigate the breach (see incident handling)

Repository Administration is managed by the GFT Project Managers (Coordinator, PM and Deputy PM) centrally. They are the ultimate responsible for the management of the Repository.

2.5.4 Encryption

All files and documents on the Repository will be kept encrypted (e.g. Google uses AES cryptography). Whenever stronger security measures will be necessary, the Folder Responsible or PSR can encrypt the files using encryption tools (e.g. zipping with password).

2.5.5 Asset Management

Asset management will involve restriction to people and organizations based on the following Information Classification (reflects the Grant Agreement and should reflect the Consortium Agreement when this will be in place):

- Public
- Restricted
- Confidential

In particular Folders and Documents will have a sharing access list based on their classification. The PSR will assure proper management of resources, administering via the Administration console.

2.5.6 Backup

All files and documents on repository have redundancy and will be automatically backed-up. Periodically (once a week) a snapshot of all the information on the repository will be downloaded and stored on GFT local file systems for backup.

2.5.7 Incident Handling

Any data breach must be reported to the PSR. The PSR may escalate to GFT Security Officers for mitigation.

2.5.8 Agnostic Solution Principle

The policies and services described above are general and agnostic of the underlined infrastructure. In principle, it should be possible to migrate the repository to any (cloud) infrastructure which exposes services for:

- File storing in separate folders
- User accounts with Logging with strong factors authentication (e.g. 2FA)
- Management of users (enable, enforce security, disable, ...)
- Management of groups (for access control)
- Archive of documents with secure encryption (natively)
- Sharing of documents with link sharing (url)
- Restrict access based on access lists
- Provide tools for editing, co-authoring
- Mailing List management
- Tools for teleconferencing

2.5.9 INFINITECH Repository

The repository of the INFINITECH Project is based on Google G-SUITE. Every partner has an account on the dedicated repository for the INFINITECH Project. An INFINITECH Project user can log in with the account company@infinitech-h2020.eu provided to all beneficiaries/organizations.

Access is granted only to infinitech-h2020.eu accounts. Access from other email and gmail.com accounts can be enabled but are **deprecated for security reasons**.

The G-SUITE drive for the INFINITECH project has been organized to store the documents and information of the project based on the structure of the Workplan (i.e. the structure reflects the work packages). The repository can also be used as a collaboration platform.

The names of the folders are prefixed by the string "`010 – FOLDER`" to have a numeric order. There are a number of folders already created to reflect the project's work plan. In particular, the repository starts at "`020 – Project`". There is a Folder created for every work package named "00X - WPX" and assigned to the Work package leader organization.

It is under the **responsibility of the Work Package Leader** to grant access and organize the information of the folders assigned.

For example, the "002-WP2" folder is under the responsibility of WP2 Leader (FTS). Access to the folders can also be regulated by mailing lists, in other words it can be accessed to all members of the group/mailing list wp2@infinitech-h2020.eu.

The repository structure is work in progress. Sub folders and permissions can be created/removed by the owner of the folder.

2.6 Data Registry

In order to keep a clear record of the data used by the project and specifically by the pilots, the following information must be collected and managed in a project data registry.

Table 2 - Record of the data used and produced in the project

Ref	Description	Type	Findable	Accessible	Interoperable	Reusable	Regulatory Compliance	Responsible	Loss/Breach/ Corruption

In particular, for the pilots' dataset, the relevant stakeholders have to report the above information required for a dataset.

The registry is kept in repository and available on line to the users via an online APP reserved to project users. Access to the Registry is made available to INFINITECH partners via secure credentials.

3 Allocation of resources

As of March 2020, the sole envisioned costs for making data FAIR in INFINITECH are possible fees to be paid to publishers to make publications open.

In this case, these fees, which are eligible expenses, will be reported as regular project expenses.

The data used/produced by the pilots will be hosted in their ISO certified testbeds under the companies' premises and policies.

As mentioned at end of the Data Security section, the sole case of long term preservation will be solved thanks to the INFINITECH OpenAIRE page that will not generate costs. Since OpenAIRE is an EU site, the length of the preservation will automatically be in line with EU directives.

4 Data security

Within Infinitech project, Data Security is achieved by managing different kind of data in different domains and scopes. The previous section [3] specifies the type of data and the specific measures to assure security.

All categories of data may be subjected to different kind of security. Risks of data should be managed with uniform and policies and strategies at all level of the project.

The following table shows the risk management for what concerns the data produced within the INFINITECH project.

Table 3 - Data Risk Management within Project

Risk	**Mitigation**	**Responsible**
Loss of data	Data will be kept in secure and redundant storage. Backups of data will be performed regularly. In case of loss of data from multiple sources and with the impossibility to recover from backups, as soon as the incident occurs, the owner of the data has to communicate it to the PM/PC and evaluate with them case by case the proper actions.	Data owner, PM and PC
Data corruption	In case of data corruption, and with the impossibility to recover from backups, as soon as the incident occurs, the owner of the data has to communicate it to the PM/PC and evaluate with them case by case the proper actions.	Data owner, PM and PC
Data breach	In case of data breach of data, as soon as the incident occurs, the owner of the data has to communicate it to the PM/PC and evaluate with them case by case the proper actions.	Data owner, PM and PC

In case of leak or loss, the data responsible has to immediately report to the Project Coordinator and Project Manager and to the Ethical Experts.

Within the scope of WP3 - BigData/IoT Data Management and Governance for SHARP Services led by LEANXCALE, there is the implementation of data governance mechanisms including eIDAS authentication, anonymization and encryption and the implementation of the regulatory compliance tools.

These mechanisms/tools will guarantee the compliance with the main legislation and directives listed in Section 2.1 as well as the pilots national rules.

The preservation and curation of the use case datasets beyond the length of the project is not required. In the case artefacts of research value will be produced by the project, they will be uploaded to the INFINITECH OpenAIRE page as explained in the Data Summary section 2 of this document.

5 Ethical aspects

The whole data management process will be supervised by the 1) Ethics Board that will be chaired by the legal and regulatory expert of the consortium (DWF GERMANY RECHTSANWALTSGESELLSCHAFT MBH) and will consist of and at least one member of the pilot sites (i.e. the pilot representatives) and of the Ethics Mentor; 2) the Ethics Mentor. Further information will be available in D2.7 (M8) and D2.8 (M15), Security and Regulatory Compliance, and in the ethics report, D10.3 : GEN - Requirement No. 3 M12 e D10.4 : GEN - Requirement No. 4 (M24).
The pilots are invited to refer to the Ethics Board and Mentor in case of any issue or doubt as well as to inform the Coordinator.

6 Conclusions

The present deliverable describes the data that will be used, produced and managed within the context of the project. Their management both for the project duration and after its end is presented, considering all the types of data that will play a role in the project, precisely: the data used and produced by the 15 pilots, possibly enriched with open data; the deliverables (confidential or public); the publications, either scientific or technical, that are based on the findings of the project; the software artefacts, released as open software or proprietary; artefacts of research value obtained from the INFINITECH infrastructure (logs or playbooks); Working documents.

In addition to the above information, the deliverable includes a presentation of the details that allow the research data to be findable, accessible, interoperable and re-usable (FAIR), with a special attention to: the handling of research data during & after the end of the project; 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).

APPENDIX A: Big Data value reference model

The following description, from the BDV SRIA European Big Data Value Strategic Research and Innovation Agenda, Version 4.0 October 2017[23] provides more information on the model taken into account for the data management in INFINITECH. The reference figure is Figure **1** - Big Data Value Chain defined by BDVA[12].

Horizontal concerns

- Data Visualisation and User Interaction: Advanced visualisation approaches for improved user experience.
- Data Analytics: Data analytics to improve data understanding, deep learning and the meaningfulness of data.
- Data Processing Architectures: Optimised and scalable architectures for analytics of both data-at-rest and data-in-motion, with low latency delivering real-time analytics.
- Data Protection: Privacy and anonymisation mechanisms to facilitate data protection. This is shown related to data management and processing as there is a strong link here, but it can also be associated with the area of cybersecurity.
- Data Management: Principles and techniques for data management.
- The Cloud and High Performance Computing (HPC): Effective Big Data processing and data management might imply the effective usage of Cloud and High Performance Computing infrastructures.
- IoT, CPS, Edge and Fog Computing: A main source of Big Data is sensor data from an IoT context and actuator interaction in Cyber Physical Systems. In order to meet real-time needs it will often be necessary to handle Big Data aspects at the edge of the system.

Vertical concerns

- Big Data Types and Semantics: The following 6 Big Data types have been identified, based on the fact that they often lead to the use of different techniques and mechanisms in the horizontal concerns, which should be considered, for instance, for data analytics and data storage: (1) Structured data; (2) Time series data; (3) Geospatial data; (4) Media, Image, Video and Audio data; (5) Text data, including Natural Language Processing data and Genomics representations; and (6) Graph data, Network/Web data and Metadata. In addition, it is important to support both the syntactical and semantic aspects of data for all Big Data types.
- Standards: Standardisation of Big Data technology areas to facilitate data integration, sharing and interoperability.
- Communication and Connectivity: Effective communication and connectivity mechanisms are necessary in providing support for Big Data.
- Cybersecurity: Big Data often need support to maintain security and trust beyond privacy and anonymisation. The aspect of trust frequently has links to trust mechanisms such as blockchain technologies, smart contracts and various forms of encryption.
- Engineering and DevOps for building Big Data Value systems.
- Marketplaces, Industrial Data Platforms and Personal Data Platforms (IDPs/PDPs), Ecosystems for Data Sharing and Innovation Support: Data platforms for data sharing include, in particular, IDPs and PDPs, but also other data sharing platforms like Research Data Platforms (RDPs) and Urban/City Data Platforms (UDPs). These platforms facilitate the efficient usage of a number of the horizontal and vertical Big Data areas, most notably data management, data processing, data protection and cybersecurity.

[23] http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf