Data Management Plan

Project Acronym: BIG

Grant Agreement No: 952226

Deliverable Number: D1.4

Corresponding Work-Package: WP1

Authors: Alexandra Mendes (alexandra.mendes@tecnico.ulisboa.pt) Nuno Nunes (nunojnunes@tecnico.ulisboa.pt), Rodrigo Rodrigues (rodrigo.miragaia.rodrigues@tecnico.ulisboa.pt)

Date: 15/09/2022

Version: 1.4

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no 952226.

Document Information

List of Contributors	
Name	Partner
Alexandra Mendes	IST / ITI/LARSyS
Nuno Nunes	IST / ITI/LARSyS
Rodrigo Rodrigues	INESC-ID / IST
Catherine Mulligan	IST

Document History			
Version	Date	Change editors	Changes
0.5	20/1/2021	Alexandra Mendes	Information to the PI's about the submission of the deliverable
0.6	19/2/2021	Alexandra Mendes	Version Ready for internal Review
0.7	20/2/2021	Nuno Nunes /Rodrigo Rodrigues	Final version for submission after internal review
0.8	22/2/2021	Nuno Nunes /Rodrigo Rodrigues	Quality check for submission
1.0	22/02/2021	Alexandra Mendes	Sign off for submission
1.0	06/09/2022	Catherine Mulligan	No input provided
1.2	12/09/2022	Nuno Nunes	Final version for submission after internal review
1.3	13/09/2022	Joana Dias	Quality check for submission
1.4	15/09/2022	Alexandra Mendes	Sign off for resubmission

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Catherine Mulligan	13/09/2022
Quality manager	Joana Dias	15/09/2022
Project Coordinator	Nuno Nunes	15/09/2022

Index

Executive Summary

BIG's Data Management Plan is produced under the scope of The European Commission Open Research Data Pilot (ORD pilot), whose goal is to grant access to the research data that is generated in H2020 projects.

The management of research data in BIG will follow the procedures and approaches set in the "Guidelines on Data Management in Horizon 2020, including preparing a detailed Data Management Plan (DMP) within the first six months of the project. BIG will adopt the best practices in research data management, following the "OECD Principles and Guidelines for Access to Research Data from Public Funding".

This document should be regarded as a living document, that will be updated throughout the project. It provides information on:

- What categories of research data will be collected, processed and/or generated by the project;
- Which methods and standards will be applied in respect of the FAIR data principles for handling research data during and after the end of the project;
- Whether research data will be shared or made publicly accessible and if so, following which open access model;
- How data will be curated and preserved, during and after the end of the project;
- Which additional safeguards will be implemented to ensure respect of the FAIR data principles in terms of allocation of resources, data security, research ethics and intellectual property rights.

As we detail below, research data that will be shared or made publicly accessible in BIG is presently anticipated to be limited in scale and scope. This is due to the objectives and nature of BIG: rather than being a purely research oriented project with mainly scientific objectives, it is a pilot project aiming to explore real-life pilots and use cases of blockchain technology and design innovation for social good. These pilots and test beds could vary from small student projects attempting to explore and test an unexpected use of blockchain technology, to an industry uses all the way to a larger testbed that could use data coming from a municipality or a governmental / non-governmental organization whose data needs to respect their own data privacy management.

Nonetheless, the project aims to evaluate on a continuous basis which research data is sufficiently anonymous, including strictly statistical and aggregate data, and this data will be made available for sharing and re-use wherever feasible.

1. Introduction

The BIG ERA Chair aims to unlock the full potential of interdisciplinary research while strengthening innovation and knowledge transfer activities in close collaboration with local and global industrial

partners and contributing to the smart specialization strategy of the Lisbon Region in stimulating the upcoming ecosystem of digital startups and their economic impact.

The notion "research data" refers to "*information, in particular facts or numbers, collected to examined and considered as a basis for reasoning, discussion or calculation*". Examples of research data include (among others) statistics, results of administrative procedures, measurements relating to exchanges, database contents, data captured in transaction logs (possibly curated to remove or reduce personal data), survey results, contents of applications, interview recordings and images. Trade secrets, commercially sensitive information and confidential information are however not considered to be research data. The implications of this scoping will be outlined in the following sections.

The BIG ERAChair on Blockchain technologies and design Innovation for social Good pursues the following strategic objectives:

- Upgrade the existing research and technological development capabilities of Instituto Superior Técnico
- Improve the innovation potential and impact of Lisbon and Portugal
- Raise international awareness about the research institutes (LARSyS and INESC-ID) and connect Técnico – Lisbon and its industry affiliates to the global knowledge networks.

One of BIG's priorities is to participate voluntarily in the Pilot on Open Research Data in Horizon 2020. This involves the delivery of a Data Management Plan within the first 6 months of the project. The Data Management Plan will detail what data the project will generate, how it will be exploited and made accessible for verification and re-use, and how it will be curated and preserved. Given the context of the BIG Project, this DMP focuses primarily on research data that are available in digital form.

2. Data Summary

Research data will be collected and processed during the BIG project for the purposes of piloting the implementation of small innovation prototypes to testbeds for decentralised, trusted, transparent, user-centric digital services and stimulating new and improved business models that promote decentralised social innovations. These testbeds will enable the industry stakeholders to test and market solutions in strategic application areas aligned with the smart specialization strategy, for instance, tourism, transportation, mobility and logistics, creative industries, and the blue economy.

BIG's data management plan's main goal is to determine how the data produced by the project will be exploited and made accessible. The data collected throughout the project in blockchain and related technologies will contribute to expanding LARSyS and INESC-ID researcher's potential and fostering a critical mass of researchers with interdisciplinary expertise.

2.1. Data sets

A large amount of data will be generated from data-centric testbeds unraveling the impact of open innovation in blockchain technologies on advanced training and knowledge transfer and expanding the human capital and the creative research capabilities of individuals and organizations with the goal of materializing opportunities offered by blockchain and distributed ledger technologies.

At this early stage of the project, three fundamental data sets were identified:

- Evidence of assets that could be used to prototype and test blockchain based services. "Evidence" in this context should be understood as any physical artifact, document, data, or digital asset that could be represented digitally through blockchain technology (e.g. an NFT representing a piece of digital art or the ownership of a physical artifact).
- Transactional data relating to exchanges: These data sets consist of the totality of data which is exchanged between parties in piloted procedures. They contain evidence as described above, but also information about the origin and the entities involved as producers or consumers of the evidence, the users involved in the procedures, and any logs or metadata relating to the characteristics of the exchanges (e.g. simulation of financial transaction or smart contracts).
- Surveys: These data sets consist of feedback obtained from any relevant user groups, including citizens, businesses and authorities which are involved (or could be involved) in testing prototypes or piloting activities.

Origin of data and re-use of existing data

As explained above, the principal data of interest in BIG is the exchanged evidence. Exploratory evidence of assets is usually collected by students and researchers from public datasets or generated from simulated data for the purpose of testing and deploying small scale innovation prototypes to test services and/or concepts for research purposes. The other two main data sets described above (transactional data and surveys) are originally generated and collected within BIG, for the purposes of organizing and evaluating the pilots, and providing recommendations in relation to the BIG/DCentral infrastructure.

Expected size of the data

The total size of the research data is difficult to estimate at this point in the project. Evidences are likely to be potentially sizable, given the fact that multiple digital entities may be exchanged, which could conceptually consist of or contain graphical imagery that substantially increases file size. Given that this data will not be eligible for sharing or re-use, this is however of limited relevance for the DMP.

Transactional data and survey data will be smaller in size, depending principally on actual piloting usage (which is not possible to predict at this stage). The amount of data generated from this analysis

is hard to predict at this stage. In general, however, it can be assumed that the expected size of these datasets will be fairly limited (< 1 gigabyte of data).

Data utility

While utility of the evidences is potentially high, these data sets are out of scope for legal reasons, and utility is thus irrelevant.

Transactional data and survey data will be useful both for policy makers and for scientific researchers. Transactional data can provide an accurate overview of piloted transactions and their main characteristics, while surveys can provide information on expectations and experiences in relation to BIG's piloting activities. As will be explained below, this potential utility must be balanced against privacy and confidentiality interests however, so that transactional data generally cannot be shared in its original form.

Therefore, the results of the project research activities will primarily be of benefit to policy makers and academic researchers. By way of example, the results may contribute to a better design of the governance model of distributed autonomous organizations (DAOs), since a better understanding of pain points and where the bigger benefits lie may allow a more optimal allocation of responsibilities.

From a more holistic point of view, the data processing and analysis performed by BIG will contribute to the well-being of the society through the subsequent streamlining and evidence of efficacy of testbeds based on blockchain technology.

Other types of data

BIG will generate other different types of data, such as:

- Articles in Scientific Journals, Conferences, and Workshops;
- Project Deliverables;
- Prototypes.

Articles in Scientific Journals, Conferences, and Workshops

All produced scientific articles will be published according to the gold" model of open access publications, which means that the articles will only be submitted to open access journals or hybrid journals that allow the open access publication through the payment of fees. Once published, the research articles will be public on the project's website. Similarly, for workshops and conferences, papers will only be submitted to conferences that allow authors to post a personal copy of the accepted papers on their websites. (Fortunately, this is the norm in top-notch computer science conferences.)

Project Deliverables

After being approved by the European Commission, each public deliverable will be made accessible on BIG's website. The list of public deliverables is the following.

WP No	Del. No	Title	Nature	Diss. Level
WP1	D1	Minutes of Kick-off meeting	Report	Public
WP1	D2	Quality Plan	Report	Public
WP1	D3	Data Management Plan	ORDP	Public
WP1	D4	Annual Report 1	Report	Public
WP1	D5	Annual Report 2	Report	Public
WP1	D6	Annual Report 3	Report	Public
WP1	D7	Annual Report 4	Report	Public
WP1	D8	Publications in high impact journals in the relevant research fields before the start date of the project	Report	Public
WP3	D11	Organization of Horizon 2020 training events v1	Report	Public
WP3	D12	Organization of Horizon 2020 training events v2	Report	Public
WP3	D13	Researcher Exchange Reports	Report	Public
WP3	D14	Strategic plan for long-lasting collaborations	Report	Public
WP3	D15	Open Research Data Pilot	ORDP	Public
WP4	D16	Characterization of Research Infrastructure	Report	Public
WP4	D17	Industry Affiliation Model	Report	Public
WP4	D18	Knowledge Transfer Plan	Report	Public
WP4	D19	Report on Knowledge Transfer and the Industry Affiliates 1	Report	Public
WP4	D20	Report on Knowledge Transfer and the Industry Affiliates 2	Report	Public
WP5	D21	Communication & Dissemination Plan – v1	Report	Public
WP5	D22	Communication & Dissemination Plan – v2	Report	Public
WP5	D23	Knowledge Management System	Report	Public

Prototypes

As a general strategy, we intend to design, build, deploy and test "in-the-wild" prototypes of data-based solutions and study their actual impact at improving global impact and citizenship. Researchers from INESC-ID and LARSyS will develop and provide open access to these prototypes. These will be demonstrated to different stakeholders such as local SMEs and startups and micro-entrepreneurs at a later stage.

The data produced by BIG will create value for the following groups:

- Industry Affiliates;
- Academic Community;
- Public Policy Community;
- Wider Public.

Industry Affiliates

By delivering strategic guidance to the community of industry affiliates, helping them to engage in new (and appropriately validated) concepts, prototypes and usages, and venues for business/academic exchange (both through established forms such as internships, seminars, or workshops, and new forms to be explored);

Academic Community

By writing new scholarly publications, frameworks and methodologies, organizing workshops or design studios, and promoting collaboration opportunities (from research-student co-supervision to co-authorship on particular projects), as well as community support (via social media channels or cross-campus initiatives);

Public Policy Community

By delivering case studies and frameworks for analysis that speaks to central issues of public concern such as sustainable behaviors, social inclusion, and public wellbeing.

Wider-Public

By delivering media outreach through press events, and written and video publications related to the real-world applications of blockchain technologies.

3. Fair Data

The BIG project attaches great importance to making its research data findable, discoverable and identifiable. Therefore following the GA and guidelines for working on documents FAIR data follows the principle that research data should be findable, accessible, interoperable, and reusable.

3.1. Findable

Data is considered findable when there's enough metadata registered or indexed in research bases to make it available to prospective users. All data that the BIG research team deems important for project participants will be integrated in the dedicated (purely internal) cloud repository suite which acts as the metasource for all project generated data. The access rights determined by the original author/publisher of included entries will be respected.

Public facing data will be disseminated via the project website, which applies standard Search Engine Optimisation (SEO) methodologies as a tool to increase the visibility and discoverability of the data based on selected keywords. SEO considers how search engines work, what users search for, the actual search terms or keywords typed into search engines and which search engines are preferred by their targeted audience. In general, via SEO the platform will appear more frequently in the search results list. SEO may also target different kinds of search, including image search, local search, video search, news search and theme-specific vertical search engines. To further improve findability, the website will likely integrate mini-sites in the future, covering specificities for individual pilots.

The project participants agree to provide adequate metadata within the data sets in order to ease the interpretation of the data and to increase the identification, discoverability, re-use and preservation thereof. Metadata is structured information describing the characteristics of the sources. A distinction is made between. BIG Researchers will be asked to provide the following metadata that describes the data produced:

- Name
- Location
- References
- Date
- Keywords
- Methods of sampling
- Equipment
- Procedures

If possible, all datasets will be documented with the above-mentioned detailed metadata. Hence, the datasets used for their publication will be publicly available for anyone to access. Other information that the research data contain include the reference period, project funding information (e.g. EU logo

and information about the Grant Agreement and the action/program that funds the project, official project name and project ID), release policy including dissemination rules, information about the collection of the data such as the data source, geographic coverage of the data, language, and file format.

Furthermore, BIG's team will publish in scientific publications such as conferences and journals. The researchers will be asked to include the acronym of the project, the grant agreement number, and a digital object identifier (DOI) in the bibliographic metadata.

3.2. Accessible

Materials generated under the BIG project will be disseminated in accordance with the Consortium Agreement. The project deliverables that are marked as 'PU' (public) in the Description of Action will be made openly available via the project website and can be further shared through related platforms such as Zenodo, OpenAIRE, etc, in accordance with the Grant Agreement and the Horizon 2020 Open Access Guide.

Certain data fall outside the scope of the open access strategy. These include different types of data that can be used to identify individuals or that are of a commercially sensitive nature. Therefore, personal data of research participants, industry partners or other stakeholders, raw qualitative research data from interviews, focus groups and workshops, draft reports, unfinished work, personal notes, plans for future research, preliminary analyses, peer reviews, and communication outside of a test setting, fall outside of the scope of the open access strategy.

Therefore, any data in Public deliverables will be anonymised. This implies that (1) evidences as a data set are entirely out of scope; (2) transactional data relating to exchanges will only be reported on at the aggregate, statistical level, except when piloting can be done through fictitious test data; and (3) survey outcomes will similarly be only reported on at the aggregate, statistical level. It is envisaged that such data will therefore not be traceable to individual users (persons, companies, government or non-government entities), nor to individual administrations or Member States when this would be reasonably likely to impair their functioning. Original (non-aggregate and thus identifiable) information will not be made openly accessible, although source information will be retained by the BIG partners for as long as legally permissible under the Consortium Agreement and/or as required under applicable law.

Nevertheles, most data that BIG will generate will be available to the broader public. These data include conference and workshop papers, journal articles, books and book chapters, deliverables (public ones) minutes of meetings, recordings of conferences and workshops, and project technical reports. This data will be accessible on the project website https://bigerachair.tecnico.ulisboa.pt/, on the project's youtube channel (when applicable), and social media.

The open research data will be made available with the lowest technical threshold possible, i.e., without any prior requirement of identification or authentication. Nonetheless, to protect the identity of research

participants and in order to encourage participants to speak freely and truthfully, all reporting and communication relating to research participants will be shared only in a pseudonymised or anonymised manner. Original (non-anonymised or non-pseudonymised data) will be stored to allow identification and traceability for research validations and follow-up, but such storage will be organized separately from the research data and in adherence to state-of-the-art confidentiality and security standards (including encryption, access logs and seals). If the document itself cannot be made secure with a password, they will be stored in an encrypted container with password protection (7zip, etc.).

3.3. Interoperable

The (meta)data that will be made open and re-usable will be in line with most widely used terminologies, standards and methodologies in the field of blockchain and DLTs, in order to facilitate interoperability and interdisciplinary interoperability.

BIG's team will follow OpenAIRE guidelines (https://guidelines.openaire.eu/en/latest/)for online interoperability.

The Project PI's will also ensure that BIG's data follows the FAIR principles under the H2020 open-access policy.
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-datamgt_en.pdf

As the project progresses and consequently more data is generated, further information on the interoperability of data will be present in the following versions of the Data Management Plan.

3.4. Reusable

The (meta)data that will be made open and re-usable will be in line with most widely used terminologies, standards and methodologies in the field of blockchain, in order to facilitate interoperability and interdisciplinary interoperability. BIG will deposit data in specialized research repositories and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate - free of charge for any user – the data (and associated metadata) needed to validate the results presented in scientific publications.

There will be no restriction on the use of data by third parties after the end of the project.

The data will remain reusable (labelled accordingly with the applicable licenses) forever.

4. Allocation of Resources

As described in the BIG Grant Agreement, within the limits of intellectual property protection and in line with the Open Research Access Pilot, BIG will disseminate the achieved results through the publication on the open section of the website of public project deliverables. Whether to use Green OA or Gold OA will be decided specifically for each research result.

As defined by the Grant Agreement, the entity responsible for data management is the project coordinator, supported by all partners who collect, generate or make data available (including notably piloting partners).

The costs for long-term preservation of the data have been considered by the consortium. There will be no further costs (other than those already provided in the general project budget) related to the preservation of data beyond the end of the funding period.

The costs related to the open-access of scientific bibliography (or participation in conferences and workshops that allow for personal copies of the papers to be posted), produced in the project's scope, are eligible for reimbursement as part of the H2020 Grant Agreement. The project Coordinator and the Research Management Unit will ensure that the Data Management Plan is followed throughout the project.

A knowledge Transfer Plan will be prepared by the ERA Chair Board, aiming to foster the transfer of knowledge generated by BIG to the broad society. This plan will also consider the overall Knowledge Transfer needs and practice of LARSyS, INESC-ID, and Técnico - Lisbon as a whole, providing an opportunity to devise a general strategy for the organization. This plan will include all the aspects relevant to Knowledge Transfer, including (i) identification of current and future areas of application of blockchain and distributed ledger technologies and design innovation approaches; (ii) identification of organizations in public and private sectors that might benefit from these approaches; (ii) development of strategies for effective knowledge transfer; (iv) exploitation of results and intellectual property rights.

5. Data Security

We will implement a Knowledge Management System (KMS), including a communication platform for disseminating and sharing knowledge, materials, methods, and results among the internal partners of Técnico - Lisbon. This intranet platform will be linked to the LARSyS and INESC-ID websites. We expect to generate a large amount of data (namely from data-centric testbeds) that will need to be made available to project collaborators. The KMS infrastructure will guarantee access to all partners and security since the platform will be protected against illicit uses (through a firewall and secure communication lines). There will also be information relevant for the management and networking in the intranet, including meeting minutes, templates, and guidelines for internal reports, progress reports, and deliverables. The intellectual property strategy will guarantee the enhanced, regulated, and sustainable use of the generated knowledge. There will be a comprehensive plan for disseminating results from BIG and the work of LARSyS and INESC-ID partners and activities described in the project. The exploitation of results will be addressed in parallel with the activation of the Living Lab. Management of Intellectual Property Rights (IPR) will be following Annex H of the H2020 Grant Agreement (GA) and regulated by the terms based upon the DESCA 2020 Model Consortium

Agreement, using the Module GOV LP for small projects. The main principles for the management of intellectual property (IP) are as follows:

- Background IP shall remain the party's property bringing it to the Project (in this case, industry stakeholders and twinning institutions). Access rights to Background IP shall be granted on the following basis: 1) for Project use ('for implementation') a non-exclusive, royalty-free license for that purpose only; 2) for exploitation ('for use') a non-exclusive, non-sublicensable license to Background IP on fair and reasonable terms shall be given from owner of the Background to another party that needs it to exploit its own Foreground IP;
- Foreground IP (results of the project) shall be owned by the party creating it. If more than one party is involved, such ownership shall be joint, in the respective proportions, and the assumption shall be that either party can exploit/'use' the joint IP provided that it gives the other notice and the other party does not object;
- Access rights to Foreground IP shall be granted as follows; on a royalty-free basis to each other party for the project's sole purpose. For exploitation, where a party needs another's Foreground to exploit its Foreground, then a non-exclusive, non-sublicensable license on royalty-free terms shall be granted;
- Standard provisions for Access Rights for any identified Affiliated Entities and IP ownership and Access Rights terms for withdrawing and joining parties shall be following standard H2020 terms incorporated within the DESCA model agreement.

6. Ethical Aspects

At the present moment, we don't foresee any ethical issues related to the project's work plan.

7. Conclusion

The purpose of this document is to set out a first, provisional version of the DMP for the BIG project. This DMP will be revised and updated throughout the entire duration of the project. Based on current assessment, research data that will be shared or made publicly accessible in BIG is anticipated to be limited in scale and scope, due to the objectives and nature of BIG.

Much of the research data in BIG is not personal data and/or sensitive in nature. The data is at any rate subject to strict legal safeguards against re-use, including through the General Data Protection Regulation and under the Single Digital Gateway Regulation, thus rendering most of the research data ineligible for sharing and re-use.

None the less, statistical and aggregate data will be generated and made available for sharing and re-use wherever possible, including specifically aggregated versions of transactional data relating to exchanges, and of surveys.

This information will be further updated throughout the project.