

## H2020-SC6-GOVERNANCE-2018-2019-2020

### DT-GOVERNANCE-05-2018-2019-2020



## D1.4 Data Management Plan

<b>Project Reference No</b>	959157 — ACROSS — H2020-SC6-GOVERNANCE-2018-2019-2020
<b>Deliverable</b>	D1.4 Data Management Plan
<b>Work package</b>	WP1: Project Management
<b>Nature</b>	Open Research Data Pilot
<b>Dissemination Level</b>	Public
<b>Date</b>	27/07/2021
<b>Status</b>	Final
<b>Editor(s)</b>	Anna Triantafyllou/ATC, Enrique Areizaga/TECNALIA
<b>Contributor(s)</b>	All partners
<b>Reviewer(s)</b>	Vincenzo Savarino (ENG)
<b>Document description</b>	Plan detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The DMP is a living document that will be modified and refined during the project period, as part of the D1.3 Project Management Report.



## About

The project is co-funded by the European Commission's Horizon 2020 research and innovation framework programme. Spanning through three years, ACROSS consists of a consortium of 10 partners from 7 countries: Athens Technology Center (coordinator), Tecniaia, Dataport, Engineering, Fraunhofer, GRNET, TimeLex, The Lisbon Council, Waag and VARAM. The project kicked off its activities in February 2021, with an energising online meeting, where all partners took the floor to present their plans to make the project a great success.

## DISCLAIMER

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use, which may be made of the information contained therein.

© 2021 – European Union. All rights reserved. Certain parts are licensed under conditions to the EU.



## Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
V0.1	14/07/2021	First draft version	TECNALIA
V0.2	21/07/2021	Comments and suggestions received by consortium partners	All
V0.3	26/07/2021	Including comment from internal reviewers	TECNALIA
V1.0	27/07/2021	Final version	TECNALIA, ATC



## Executive Summary

This deliverable, D1.4 Data Management Plan (DMP), presents the Plan detailing what data the ACROSS project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. During the course of the project, ACROSS will generate data in a wide range of activities. Since the ACROSS project is at an early stage, it is important that timely dissemination of these findings (data, publications, survey results) are open for scrutiny by other researchers, potential future partners, and the wider research community.

As a project participating in the Open Research Data Pilot (ORDP) in H2020, ACROSS will make its research data findable, accessible, interoperable, and reusable (FAIR). Nevertheless, data sharing in the open domain can be restricted, taking in account “the need to balance openness and protection of scientific information, and privacy concerns, security as well as data management and preservation questions” as stated in Guidelines on FAIR Data Management in Horizon 2020 published by the European Commission.

The DMP’s purpose is, therefore, to provide the main elements of the data management. To this end, it describes types and formats of data to be generated or collected and how the standards to be applied, the data-reservation methods, the data sharing policies for re-use. The present document is the first version of the ACROSS DMP, containing an initial overview of the foreseen ACROSS datasets. The DMP is a living document that will be modified and refined during the project period, as part of the D1.3 Project Management Report.



## Table of Contents

<b>1</b>	<b>INTRODUCTION</b> .....	<b>1</b>
1.1	PURPOSE AND SCOPE .....	1
1.2	APPROACH FOR WORK PACKAGE AND RELATION TO OTHER WORK PACKAGES AND DELIVERABLES .....	1
1.3	METHODOLOGY AND STRUCTURE OF THE DELIVERABLE .....	1
<b>2</b>	<b>DATA SUMMARY</b> .....	<b>2</b>
2.1	DATA RELATED TO THE USE CASES .....	2
2.1.1	<i>Greece</i> .....	4
2.1.2	<i>Germany</i> .....	4
2.1.3	<i>Latvia</i> .....	5
2.2	DATA RELATED TO SCIENTIFIC PUBLICATIONS .....	5
2.3	ACROSS PUBLIC DELIVERABLES .....	6
<b>3</b>	<b>FAIR DATA</b> .....	<b>7</b>
3.1	MAKING DATA FINDABLE .....	7
3.2	MAKING DATA OPENLY ACCESSIBLE .....	8
3.3	MAKING DATA INTEROPERABLE .....	9
3.4	INCREASE DATA RE-USE .....	9
<b>4</b>	<b>ALLOCATION OF RESOURCES</b> .....	<b>10</b>
<b>5</b>	<b>DATA SECURITY</b> .....	<b>11</b>
<b>6</b>	<b>ETHICAL ASPECTS</b> .....	<b>13</b>
<b>7</b>	<b>OTHER</b> .....	<b>13</b>
<b>8</b>	<b>CONCLUSIONS</b> .....	<b>14</b>
<b>9</b>	<b>REFERENCES</b> .....	<b>15</b>



## List of Figures

FIGURE 1 - ACROSS HIGH-LEVEL ARCHITECTURE (SOURCE: DOA).....	11
FIGURE 2 - THE ENVIRONMENTS IN ACROSS.....	12

## List of Terms and Abbreviations

Abbreviation	Definition
EC	European Commission
BPM	Business Process Model
BPMN	Business Process Model Notation
CAS	Central Authentication service
CPSV	Common Public Service Vocabulary
CPV	Common Person Vocabulary
DCAT-AP	Data Catalogue vocabulary – Application Profile
DMP	Data Management Plan
DoA	Description of action
DPIA	Data protection impact assessment
GDPR	General Data Protection Regulation
IPR	Intellectual property rights
ISP	Innovation Support Platform
LDAP	Lightweight Directory Access Protocol
PA	Public Administrations
PIA	Privacy impact assessment
WP	Work package
Yx	Year x, where x is a number



## 1 Introduction

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable, and re-usable (FAIR), a DMP should include information on:

- 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 Purpose and Scope

This deliverable focuses on the management of the data in ACROSS. In ACROSS there will be two different data, the first strand related to the publications generated as part of the research activities, and the second strand related to the data collected from stakeholders (users and also service providers) will be aggregated into data sets that will comprise statistical and therefore anonymous data and will be used as part of the implementation of the different key results established in the project.

### 1.2 Approach for Work Package and Relation to other Work Packages and Deliverables

This deliverable of the ACROSS project is prepared under WP1 and the Task 1.1. In this Task we initiate discussion of the data management processes and/or data generated by the ACROSS project in order to make the data findable, accessible, interoperable and reusable (FAIR). This data management plan is a living document that will be edited and updated during the project period, with a second version to be delivered in month M12 as part of the D1.3 Project Management Report.

### 1.3 Methodology and Structure of the Deliverable

The document follows the established H2020 template for a Data Management Plan (DMP) [1] and is comprised of the following Chapters:

- Section 2 presents a summary of what the purpose of the data collection and generation is in the case of ACROSS.
- Section 3 explains how the data and metadata will be made accessible, findable, and reusable.
- Section 4 briefly explains how the financial resources for this openness are envisioned at this stage to be allocated while Sections 5 and 6 focus on the security and ethical aspects respectively.



## 2 Data Summary

ACROSS aims to enhance the use of cross-border public digital services during work- or study-related moves, producing an integrated eco-system that truly places the citizens at its core. Interaction with these citizens (or end users) is therefore crucial to the project.

During the initial months of the project, the inputs of end users will be done in the form of co-creation sessions. These can take many forms, including direct individual interviews, workshops, interactive discussion sessions, focus groups, and so forth. In all these cases, direct inputs from the participants in relation to their needs, preferences and expectations are sought through direct interaction.

This interaction will be done only after the signing of informed consent forms and mainly in the form of (individual or focus group) interviews, the outcomes of which will serve to formulate user journeys. Since the interest of the project is primarily in trends rather than in individual responses, the feedback from individual users (which can be attributed directly to them and therefore constitutes personal data) will be aggregated into data sets that will comprise statistical and therefore anonymous data. That anonymous data will then be translated into personas, so that they are not traceable to individuals (and therefore no longer constitute personal data) yet serve as representation of real-life experience.

These personas will be the basis for piloting activities in the later stages of the project. As a result, piloting outcomes similarly will not be based on real persons (although the personas will be credible and realistic to ensure usable project outcomes), and therefore piloting results also will not qualify as personal data.

The first phase of interaction with users will be done during M6 and M7 and will consist of a minimum of ten interviewees per use case. In later phases, interaction will also be done with other stakeholders, including public administrators, private companies, employers, and educational institutions. The aggregated and anonymized data from this interaction will further help the formulation of user journeys (Task 2.1) and personas, will provide insights into current gaps in service provision (Task 2.2), and will establish a basis for the refinement of the technical requirements of the ACROSS platform.

Evaluation of the ACROSS results might imply involving “real personas” (citizens outside consortia partners). Should this be the case, their personal data will be anonymized.

### 2.1 Data related to the use cases

The main purpose of the data collection in the three use case sites is to analyze the current situation with cross-border access to services by identifying existing problems and their possible solutions in a user friendly way by gaining insight from the service providers, on one hand, and service recipients (end-users), on the other hand.





The project aims to improve the use of cross border digital public services. This requires a good understanding of the preferences, needs and expectations of the stakeholders, principally users but potentially also service providers. For that reason, real life data must be collected.

The final outcomes of the co-creation sessions will be textual and numerical data. Possibly video and audio recordings will be used in some of the sessions, but these are in principle intermediary data that only serve to create the textual and numerical data (i.e. recordings principally serve to produce transcripts and to verify impressions; they have no independent use apart from serving as a source to product robust textual and numerical data).

The current approach is that recording (either audio or video) is permissible for internal project purposes only - so, for transcribing answers, analysing responses, detecting patterns etc; but not for public dissemination. There are transparency and consent notices in that sense “Detailed variant for group interviews - combined informed consent / information notice for internal use only (no publication)”.

There’s also a separate template in cases where ACROSS would like to publish recordings, e.g. on the project website or social media; that’s in the same deliverable under the heading Simple variant - combined informed consent / information notice via e-mail, with responses intended for internal use and for publication on project website / social media channels. Those should be used only for individual interviews with experts.

Recordings will be saved only within the closed Teams environment and only during the timespan of the project would suffice in terms of security.

Statistical and aggregate (therefore non-personal) data related to the provision of public services that is collected on regular bases by the service owners (Pas) will be re-used.

Data related to end-users experience and opinions will be gathered from end-users themselves during interviews, focus group discussions or other qualitative research methods.

Data related to public service usage will be gathered from the public institutions.

The expected size of the data is not known at the moment, most likely for each use case it would be under 50 GB.

It is presently envisaged that the data will only be made available to project partners that require it for the purposes of the project, specifically to create statistical and aggregate data and to define personas. No further use is planned, although the data conceptually may be useful to scientific researchers interested in validating or repeating the research.



### 2.1.1 Greece

For the Greek use case, a minimum of ten interviews will be conducted with EU member states' citizens who moved recently to another EU country (any country including Great Britain) for work- or study-related reasons. At least five one-on-one interviews per 'user profile' (work-related move vs. study-related move) are desired and the questions will be open, giving the chance to the interviewee to express his/her experience freely.

The form of the interview is based on the common format which was decided by the ACROSS project partners and includes questions regarding demographic data, the scenario of the move and qualitative information about the moving process, relevance with specific cross-border services, gaps between services, recommendations and data privacy issues.

The results will be anonymous and labeled. Names or other personal characteristics are not reused. The interviews will be recorded, after the consent of the interviewee, and the recording will only be used for the depiction of his/hers journey on a template. After that the recording will be destroyed.

The aim of this research is to provide the technical team with potential user journeys in order to set up the requirements, attributes and boundaries of the ACROSS platform modules.

### 2.1.2 Germany

The aim of the user research process for the German use case is – the same applies to the other use case owners - to form potential user journeys, reusable for the technical team in order to set up the requirements, attributes and boundaries of the ACROSS platform modules. Hence, these user journeys can include actions, touch points, emotions, pain points, and phases.

Even though, various forms of interaction are possible, Germany decides to focus on qualitative research, primarily one-on-one interviews. At least five one-on-one interviews per 'user profile' (work-related move vs. study-related move) are desired. This minimum of ten conducted interviews will then be distilled into various personas at a later stage (refer to Personas WP2 - WP6 document for elaboration). The interviewees can be moving from or to any European country (not limited to moves between Germany, Greece, and Latvia). It is recommended, if possible, to interview people who moved both to and away from their own pilot country.

The conduction of the interviews mostly consists of open questions. Each interview participant can answer freely. No one is steered in a particular direction. No one is influenced in any way to achieve a particular result that the interview partner would not want to talk about.



In addition, the results are anonymous and labeled. Names or other personal characteristics are not reused. Notes will be taken, but no transcription or recording will be done.

In case, interviews do not deliver the envisioned output (dealing with the ACROSS user journey), the use case owner Dataport from Germany potentially conducts surveys, focus groups or co-creative sessions – depending on the capacity.

### 2.1.3 Latvia

Similarly to previously described use cases of Greece and Germany, Latvia will collect data from qualitative research by interviewing end-users of cross-border public services about their experience in work or study related moves within different EU countries. No sensitive information will be collected. Personal information (i.e. the responses that can be attributed to individual participants) will not be disclosed to parties outside of the consortium or made public and will be managed according to respective legal framework.

In addition to qualitative research of end-users' experience, statistical data and other information about relevant public services, e.g., compliance with the requirements of the SDG Regulation and corresponding national regulatory framework, assessment of the digital maturity of public services, number of service users, etc.) will be collected from other national regulatory authorities. This will include publicly available open data published on Latvia's Open Data portal <https://data.gov.lv/eng> and data collected by the owners of said public services – public institutions. Types of data will be in widely used formats, as MS Word (.doc/.docx), MS Excel (.xls/.xlsx), PDF (.pdf), XHTML or HTML (.xhtml/.htm). No personal or sensitive data will be collected from national regulatory authorities.

## 2.2 Data related to scientific publications

According to the European Commission, “under Horizon 2020, each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results” (see also Article 29.2 of the GA). The ACROSS Consortium adheres to the EU open access to publications policy, choosing as most appropriate route towards open access self-archiving (hereinafter also referred to as 'green' open access), namely “a published article or the final peer-reviewed manuscript is archived (deposited) in an online repository before, alongside or after its publication. Repository software usually allows authors to delay access to the article ('embargo period)’”.

This dataset will contain manuscripts reporting the conducted scientific work in ACROSS which have been accepted for publication in peer-reviewed journals and conferences. Such data may (and likely will) contain aggregate and statistical data, and describe the personas created in ACROSS, but no personal data



from the co-creation sessions will be included. All these publications will include a statement with acknowledgement to the ACROSS project while their content may vary from the description of specific analysis techniques, to established evaluation datasets and individual components or parts of the ACROSS platform.

Most commonly, these documents will be stored in PDF format. Each document will be also accompanied by: (a) details about the venue (e.g. conference, workshop or benchmarking activity) or journal where it was published, (b) a short description with the abstract of the publications, and (c) the LaTeX-related BIB file with its citation. This dataset will be extended whenever new submitted works are accepted for publication in conferences or journals.

This dataset will be publicly available, following the guidelines of the EC for open access to scientific publications and research data in Horizon2020 [2].

Self-archiving (also known as "green" open access) will be applied for ensuring open access to these publications. According to this archiving policy the author(s) of the publication will archive (deposit) the published article or the final peer-reviewed manuscript in online repositories, such as personal webpage(s), the project website and the free-of-charge OpenAIRE or Zenodo [2] repositories, after its publication. Nevertheless, the employed archiving policy will also be fully aligned with restrictions concerning embargo periods that may be defined by the publishers of these publications, making the latter publicly available in certain repositories only after their embargo period has elapsed.

### 2.3 ACROSS public deliverables

ACROSS deliverables will be stored in PDF format. For each deliverable we will provide: (a) the list of authors, (b) a brief description of its content (i.e. its abstract), (c) the related WP of the project, and (d) the contractual date for their submission to the EC. This dataset will be extended whenever new deliverables are submitted to the EC. As with scientific publications, the deliverables can contain aggregate and statistical data, and describe the personas, but no personal data from the co-creation sessions will be included. The public project deliverables will be made publicly available after their submission to the EC, via the project website <https://across-h2020.eu/>. This dataset will be maintained on the relevant webpage of the project website. This webpage will grant open access to the PDF file of each listed public deliverable once it has been approved as accepted by the PO at the end of the reporting period.



## 3 Fair Data

### 3.1 Making data findable

The ACROSS project attaches great importance to making its research data findable, discoverable and identifiable. Following the consortium agreement and guidelines for working on documents, each version of a document is identified at the beginning of the document. Naming conventions were communicated to the project partners during the kick-off meeting.

All data that the consortium deems important for all project participants will be integrated in the internal project repository, a Teams based storage [4] space which acts as the general source for all project generated data that must be accessible to multiple partners.

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 partners will provide adequate metadata for relevant 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:

- Descriptive metadata, such as title, abstract, author, and keywords,
- Administrative metadata which are used to provide information to help manage a source, such as when and how it was created, file type and other technical information, and who can access it.

In order to increase the findability, the partners will also include keywords or key-phrases describing the subject or content of the data including relevant terms of the field

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.



A unique identifier will be assigned to each dataset. In the spirit of linked data these will ideally be URIs. Depending on the formats chosen, timestamps will also be assigned.

### 3.2 Making data openly accessible

Materials generated under the ACROSS 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 [2], 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, including principally co-creation outputs (specifically the raw data from these sessions). As a consequence, personal data of research participants, project 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 PU deliverables will be anonymised. This implies that co-creation responses as a data set (including any audio and video recordings) are entirely out of scope; and that survey outcomes will be only reported on at the aggregate, statistical level. It is envisaged that such data will therefore not be traceable to individual users (persons or companies), nor to individual administrations, companies 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 ACROSS partners for as long as legally permissible under the Consortium Agreement and/or as required under applicable law.

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, in order 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 in order to allow identification and traceability for research validations and follow-up, but such storage will be organised separately from the research data and in adherence to suitable confidentiality and security standards.



Moreover, depending on the licensing by the data source maintainers there may be restrictions on what data can be published openly. The primary way of access to the data will be via HTTP REST API. The APIs will be specified using the OpenAPI standard. The data and metadata will be stored in their respective databases. Where and how the documentation will be published has yet to be determined. The code will be available in the official ACROSS Gitlab.

### 3.3 Making data interoperable

The metadata structure will cohere to the DCAT-AP specification. Data will be transformed into a domain specific standard, if applicable. For example, Datex II is one of the established standard formats for traffic flow related data. The final list of these standards has yet to be determined.

### 3.4 Increase data re-use

Specific details on future use of the data will be provided in the sustainability task (notably Task 7.4. IPR Management, Exploitation Strategy and Sustainability). Where feasible, the project consortium aims to apply open licensing through common, standardised and widely known license models, such as e.g. a 'CC-BY-SA 3.0' license, as a general rule to all research data in order to facilitate the widest re-use as possible, or e.g. the EUPL for software releases.

All open research data will be made available for re-use without any data embargo, meaning that all data will be made openly available and free to re-use upon their publication.

Furthermore definitive decisions in relation to the publishing/sharing of data with regards to licenses and point in time can only be decided upon once the selection of data sources has been finalized and thoroughly evaluated.



## 4 Allocation of resources

During the project, the data will be stored in TECNALIA's High performance infrastructure at no cost for the consortium.

In the case of open source software, the TECNALIA will ensure that the GitLab repository is available after the project duration, either by keeping it in its own premises or by transferring it to existing open source projects and communities.

No additional allocation of resources, beyond the ones already participating in the project, is expected.



## 5 Data Security

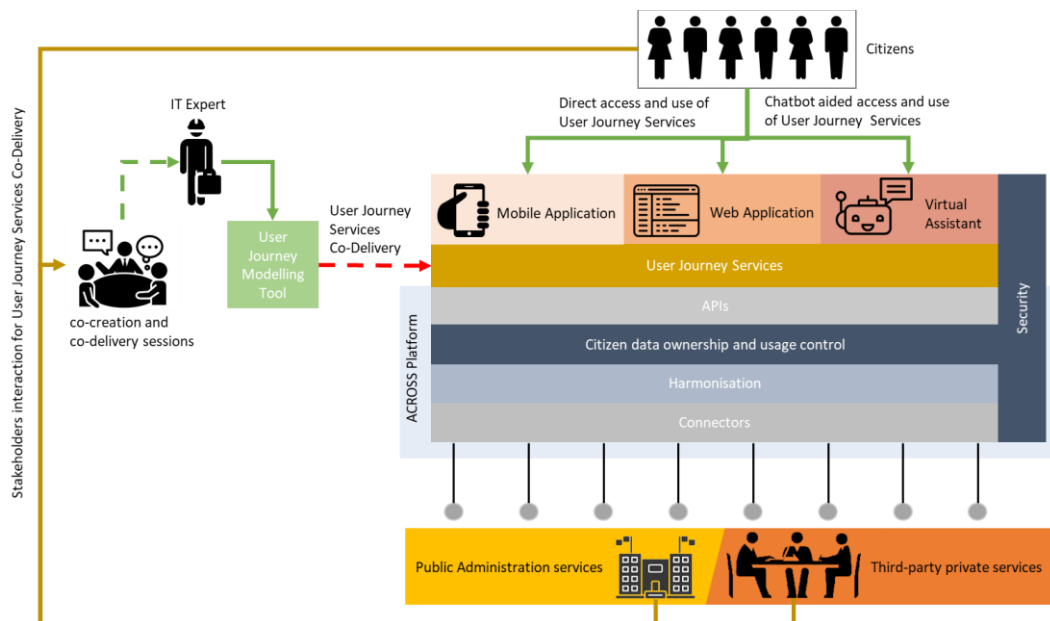
**Address data recovery as well as secure storage and transfer of sensitive data.**

The ACROSS ecosystem will adhere to security – by – design and privacy – by design principles that allow the security and audit standards to remain consistent across multiple environments.

Furthermore, ACROSS envisions the following security – related aspects:

1. Access Management, which grant authorized users the right to use a service, while preventing access to non-authorized users of ACROSS;
2. Credentials Management to manage credential information such as usernames and passwords, and
3. Anonymization, which is responsible for removing personally identifiable information from user data or pseudonymizing the personally identifiable information.

Furthermore, and while the architecture will be reviewed in the deliverable D5.1 [2], a security component was already foreseen in the high-level architecture presented in the DoA [3], and copied below:



**Figure 1 - ACROSS high-level architecture (Source: DOA)**

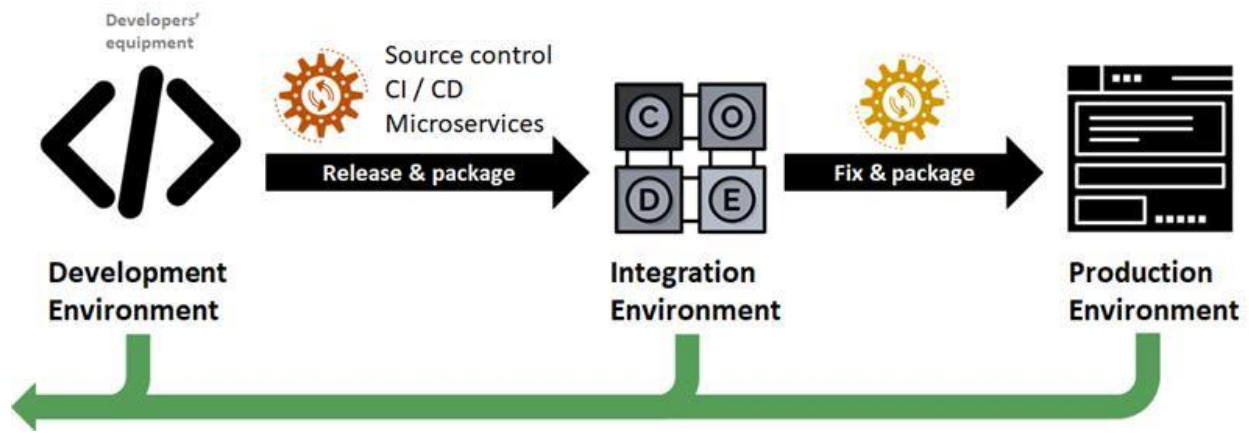


Figure 2 - The environments in ACROSS

At development stage each partner manages the security of their development environments. The source code will be uploaded to GitLab, hosted in TECNALIA, which follows the LDAP [7] authentication protocol. The integration environment will include a working version of the ACROSS ecosystem, implementing the functionalities already explained above of access and credentials management. The third environment is the actual deployment of the different components of ACROSS in the use cases. ACROSS follows a modular approach, so not all components will have to be installed in the use cases, just those that are of interest for the pilot cities. The security in this case shall follow the policies defined in the different municipalities.



## 6 Ethical Aspects

The ethical aspects of the ACROSS project will be assessed under Work Package 8, which sets out the ethics requirements that the project must comply with, and includes 14 separate deliverables, each addressing a specific topic.

Key obligations and constraints assumed by the project include:

- The appointment of a Data Protection Officer (DPO), and additionally of an Ethics Advisor;
- The definition of standardized procedures and criteria that will be used to identify/recruit research participants;
- The definition of standardized information and consent templates and procedures for any data collection sessions;
- The definition of data collection constraints, notably to avoid collecting sensitive data or data in relation to persons in a vulnerable situation;
- A description of relevant security and confidentiality measures;
- The definition of an anonymization and pseudonymization strategy;
- Explicit confirmation that the beneficiary has lawful basis for the data processing, and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects;
- A general assessment of ethics risks other than data protection challenges.

Moreover, as already highlighted above, the project applied a data protection by design approach during the proposal stage already, through the use of personas for piloting purposes. This allows detailed data collection and analysis, based on realistic user profiles, without running any data protection or privacy risks.

## 7 Other

To be updated in future versions with other national/funder/sectorial/departmental procedures for data management



## 8 Conclusions

The deliverable at hand has presented an initial version of the plan for the data management in the ACROSS project. In this action, different types of data are envisioned to be collected and generated: data coming from the use cases, related to the data management platform, from the publications, open source software and the deliverables.

The data used will either be anonymous, and from open data repositories, or sought to be anonymized as much as possible. Whenever personal data is collected, explicit consent will be required, and data will be thereafter anonymized as soon as reasonably practicable.

All data and metadata generated will follow machine-readable formats. The exact format cannot be explicated at this stage.

Data from the use cases will be stored in the use cases' premises fulfilling the relevant legislations. Data from publications will be stored in OpenAire indexed repositories favouring the green model whenever possible. Other publications such as deliverables will be stored at TECNALIA's hosting services.

This deliverable will be updated in subsequent releases, namely in M12, M24 and M36 as part of the Project Management reports. It is envisioned that in those versions the aspects that at this stage are not fully clear (e.g. naming conventions, versioning, dataset size) will be clarified as work progresses in all the work packages.



## 9 References

- [1] European Commission, “Data Management Template Annotated,” [Online]. Available: [https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf) . [Accessed 2020].
- [2] European Commission, «Guidelines on Open Access to Scientific Publications and Research Data,» 2017 March. [En línea]. Available: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- [3] Zenodo, «Zenodo,» June 2020. [En línea]. Available: <http://www.zenodo.org>
- [4] «Team based storage space,» [En línea]. Available: <https://www.microsoft.com/en-ww/microsoft-teams/group-chat-software>.
- [5] ACROSS Consortium, «D5.1 System architecture & Implementation,» 2021
- [6] ACROSS Consortium, “Description of Action,» 2021.
- [7] «Lighthweight Directory Access Protocol,» [En línea]. Available: <https://ldap.com/>.