

H2020-SC6-GOVERNANCE-2018-2019-2020

DT-GOVERNANCE-05-2018-2019-2020



D1.4 Data Management Plan

Project Reference No	959157 — ACROSS — H2020-SC6-GOVERNANCE-2018-2019-2020
Deliverable	D1.4 Data Management Plan
Work package	WP1: Project Management
Nature	Open Research Data Pilot
Dissemination Level	Public
Date	30.04.2024
Status	Final
Editor(s)	Enrique Areizaga (TECNALIA)
Contributor(s)	All partners
Reviewer(s)	Vincenzo Savarino (ENG)
Document description	The plan outlines the data that the project has generated, detailing its exploitation, accessibility for verification and re-use, as well as its curation and preservation methods. The Data Management Plan (DMP) has been modified and refined throughout the project period.



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), Tecnalia, Dataport, Engineering, Fraunhofer, GRNET, TimeLex, The Lisbon Council, Waag and VARAM.

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.

© 2024 – 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 (1 st version)	TECNALIA, ATC
V1.1	19/4/2024	Updated version based on partners input	TECNALIA
V1.2	22/4/2024	Internal Review	ENG
V2.0	30/04/2024	Final version updated.	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 project, ACROSS has generated data in a wide range of activities. ACROSS project ensured that 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 has made 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 final version of the ACROSS DMP, containing ACROSS datasets. The DMP has been a living document that was modified and refined during the project period, as part of the D1.3 Project Management Report.

In its final version all updates collected from the partners reflect the real implementation of the DMP.



Table of Contents

1	INTRODUCTION	1
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
2	DATA SUMMARY	2
2.1	DATA RELATED TO THE USE CASES	3
2.1.1	<i>Greece</i>	5
2.1.2	<i>Germany</i>	8
2.1.3	<i>Latvia</i>	10
2.2	DATA RELATED TO SCIENTIFIC PUBLICATIONS	12
2.3	ACROSS PUBLIC DELIVERABLES	13
3	FAIR DATA	14
3.1	MAKING DATA FINDABLE	14
3.2	MAKING DATA OPENLY ACCESSIBLE	15
3.3	MAKING DATA INTEROPERABLE	16
3.4	INCREASE DATA RE-USE	16
4	ALLOCATION OF RESOURCES	17
5	DATA SECURITY	18
6	ETHICAL ASPECTS	20
7	CONCLUSIONS	21
8	REFERENCES	22



List of Figures

FIGURE 1 - ACROSS HIGH-LEVEL ARCHITECTURE (SOURCE: DOA).....	18
FIGURE 2 - THE ENVIRONMENTS IN ACROSS.....	19

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 have been 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). They have been aggregated into data sets that comprise statistical and therefore anonymous data and have been 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 has been edited and updated during the project period, with a final version delivered in month M39.

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.



2 Data Summary

ACROSS enhances 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) has been therefore crucial to the project.

Data collection for the pilot use cases in Greece, Germany and Latvia occurred in three phases. These phases comprised a research process to define the use case scenarios and thereby the user journeys (see D6.1), an initial Use Case Evaluation and Impact Assessment approach (D6.2), and a final Use Case Evaluation and Impact Assessment approach (D6.3). All research conducted by pilot partners (GRNET, DATAPORT, VARAM) was supported by our legal and ethics partner (TIMELEX) as well as by WAAG in applying the User Journey methodology and *Lisbon Council* for designing the evaluation framework, creating the data collection instruments, and analyzing the received feedback.

Non-personally identifiable user data were collected in various ways during these three data collection phases.

During the initial months of the project, the input of end users was done in the form of co-creation sessions. These were done in many forms, including direct individual interviews, workshops, and surveys. In all these cases, direct input from the participants in relation to their needs, preferences and expectations were sought through direct interaction.

Further data was collected for an initial evaluation and the final impact assessment of ACROSS platform from the three use cases. Initial usability tests took place in the three pilot countries for the alpha version of the platform with the use of anonymous online questionnaires. Similarly, the impact assessment performed on the final version of the platform required online anonymous surveys to be filled in by the stakeholders where they could provide their feedback on the platform in a structured questionnaire.

This interaction was done only after the signing of informed consent forms and mainly in the form of (individual or focus group) interviews, the outcomes of which served to formulate user journeys. The informed consent forms templates and information sheets have been created by the legal and ethics partner (Timelex) and are part of the Ethics Deliverables. Since the interest of the project was 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) was aggregated into data sets that comprised statistical and therefore anonymous data.



2.1 Data related to the use cases

The main purposes of the data collection in the three use cases were on the one hand 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, and service recipients (end-users) and on the other hand to assess the proposed solution.

The project aims to improve the use of cross border digital public services. This required 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 was collected. The final outcomes of the co-creation sessions and surveys were textual and numerical data.

Video and audio recordings were used in some of the sessions, but these were in principle intermediary data that only serve to create the textual and numerical data (i.e. recordings principally served to produce transcripts and to verify impressions; they had no independent use apart from serving as a source to product robust textual and numerical data). The approach was that recording (either audio or video) was permissible for internal project purposes only - so, for transcribing answers, analyzing responses, detecting patterns etc; but not for public dissemination. There are transparency notices and informed consent forms notices in that sense “Detailed variant for group interviews - combined informed consent / information notice for internal use only (no publication)”. This template can be found in D8.2 (initial templates for informed consent forms and information sheets). 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 (D8.2) 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 were saved only within the closed Teams environment and only during the timespan of the project would suffice in terms of security.

Data related to end-users' experience and opinions was gathered from end-users themselves during interviews, focus group discussions, surveys or other qualitative research methods. The partners remained prudent that they did not collect more data than necessary, or sensitive data (i.e. data concerning health or race).

Data related to public service usage was gathered from public institutions. Statistical and aggregate (therefore non-personal) data related to the provision of public services that was collected on a regular basis by the service owners (Pas) was re-used. The size of the data for each use case is under 1 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. No further use is planned, although the data conceptually may be useful to scientific researchers interested in validating or repeating the research. This intended further use is included in the informed consent forms and the information sheets, so the research participants and stakeholders are informed.

During the impact assessment phase of the project, ACROSS survey data was downloaded from the EU Survey Tool in form of Excel files (.xlsx) which were stored on partners' institutional computers, the shared consortium Teams managed by ATC and the Lisbon Council (LC) internal Google Drive. All of these storage solutions are password protected and accessible only by team members involved in the ACROSS project.

It should be highlighted again that the survey data (i.e. the answers of the participants to the survey) itself does not contain any personal data, as this data was collected in a fully anonymized manner. However, in order to contact the research participants to participate in the survey, personal data was collected. Moreover, certain answers such as their opinions around the platform or their experiences with moving for work/studies could (potentially) reveal their identity. Therefore, the Consortium has opted for collecting informed consents from all of the participants before the (anonymized) data collection through the ACROSS survey.

In response to privacy concerns associated with cloud storage services based outside the European Union (such as Google Drive), stakeholder personal data will be stored in an alternative document repository that stores data in the EU and respects GDPR (ATC and LC's internal servers). Personal data from stakeholders will only be kept for the time necessary for the ACROSS consortium to complete research and other project activities and follow-up, including evaluation. All such information and personal data will be destroyed within 5 years after the completion of the project, i.e. by April 2029, in accordance with the data retention information included in the informed consent forms and information sheets that have been provided to the research participants. Personal data from stakeholders will not be shared among partners or with any third parties.

Non-personal, fully anonymized information and data from ACROSS is shared within the consortium via the dedicated ACROSS Teams, both for the purposes of ACROSS research and activities, and for future research or follow-up projects if relevant. The anonymized data coming from the surveys is solely used to:

- Draw relevant conclusions for the project.
- Create anonymous statistical information used for impact assessment and/or other relevant publications, without disclosing the identity of respondents.
- Ensuring the scientific validity and security of the data collection and analysis.



Participants to the impact assessment that took the survey were asked by the Pilot partners to read and sign the informed consent prior to taking the survey, explicitly acknowledging their voluntary participation and the collection of personal data. The informed consent was provided to the participants either in person on a printed document, personally signed upon consent, or electronically via email and returned with a handwritten signature. It was made clear that participation was entirely voluntary, and that the participants were free to withdraw and remove their data from the study at any time, if they so wish. All participants were asked to sign an “informed consent form” that:

- are in a language and in terms fully understandable to them.
- describe the aims, methods and implications of the research, the nature of the participation and any benefits, risks or discomfort that might be involved.
- contains all information as required by Article 14 of the GDPR and should be considered a compliant consent in the sense of Article 7 of the GDPR.
- explicitly state that participation is voluntary and that anyone has the right to refuse to participate and to withdraw their consent for participation and for data collection and use, samples or data at any time without any consequences.

The following subsections describe in more detail the data collection process per pilot use case in Greece, Germany and Latvia. More details about each of the three phases, the collected data and how it has been processed can be found in Deliverables D6.1, D2.1, D6.2 and D6.3, respectively.

2.1.1 Greece

1. Use Case Scenarios & Roadmap:

In the initial phase of ACROSS and in order to determine which direction the project would take, the consortium first had to map the current experiences and services used by the EU citizens who move abroad for working and studying. To do so, a co-creative process started in WP2 with the aim to make clear the user journeys of citizens who move across Europe to work or study. Pilot partners conducted interviews, to support this process, with people in Greece, Latvia, and Germany who had crossed borders for work and/or study purposes in the past.

Based on a ‘starter pack’ for interviews (questionnaire, consent form etc) which was created by WAAG and other partners (see D2.1, D6.1) to help pilots embark on the interview and survey process, GRNET carried out interviews over the summer of 2021 (M7) with 10 people who actually moved across European borders for work or study within the last 5 years. This was the main input for the user scenarios. In a second co-creation workshop, the WP2 and WP6 partners came together to collect and



compile the content from the interviews. The outcome of this process was the development of two shared user scenarios that are applicable to the three pilot countries: one for working abroad and one for studying abroad (See D6.1).

GRNET team collected data privacy agreements from all participants. Interviews were recorded temporarily, given the consent from the interviewees, as a backup for ensuring the accuracy of the notes and that all the necessary information would be included. The given interview template was slightly adjusted due to different situations and approaches. The approach used during interviews was that of storytelling and had a positive impact on the result, since letting the interviewees tell their story of moving abroad through their own eyes revealed many positive and negative points that were directly linked to the interview questions.

2. Use Case Evaluation and Impact Assessment – Initial:

In Greece, the usability test of the Alpha version of the ACROSS platform took place during July 2022 (M19). The test conducted examined the usability of the across alpha version in terms of design, understandability, emotions, data security and authentication (See D6.2).

Twelve Greek citizens navigated in the platform and filled in an online questionnaire with questions about themselves as well as their opinion about the platform. The questionnaire that was used was common for all pilots, for the comparability of the results. In total there were 30 questions. All 12 participants had previous experience of moving abroad for work or studies with the majority having moved after 2015.

There were two ways to participate to the survey, either by 30-minute personal online interviews, where the interviewer guided the conversation, or independently by following the instructions in the questionnaire and answering the questions. Interviewees were informed about the usability tests via invitation, which also included information about the use of their personal data: these were anonymised, and they were collected only with their consent and could be withdrawn at any moment.

3. Use Case Evaluation and Impact Assessment - Final:

To measure the impact of the ACROSS solution (second version) as applied in the three pilots, D6.2 had defined the impact assessment framework and methodology, as well as the instruments for the data collection of the applied pilot solutions, concluding in a set of 47 quantitative and qualitative indicators to measure the effectiveness, efficiency, relevance, replicability and scalability of the ACROSS intervention in the three pilots for citizens, public administrations and business. Based on



these indicators, a set of assessment materials was developed in WP6 to support three extensive surveys conducted in each pilot country addressed to both citizens and experts (policy makers, service providers, public and private sector officials, IT experts). Survey material includes two online questionnaires, a presentation, usability test instructions, consent forms etc. Pilot partners (GRNET DATAPORT, VARAM) and Lisbon Council prepared the material and deployed the questionnaires in the EU Survey Portal as anonymous surveys. The Partners have also been constantly in touch with LC to report any possible issues or malfunctions of the tools during the assessment activities.

In the surveys conducted between M34 and M37, 45 citizens and 13 experts participated in Greece; above the target that was set (See D6.3). GRNET defined a calendar of campaigns and in-presence activities in order to follow the plan for collecting the target surveys until the established deadline. For this purpose, GRNET prepared email invitations that were sent to mailing lists and candidate participants asking them for their consent to the consent form (see D6.3 APPENDIX 4). Consent Form and Privacy Note were prepared by Timelex and were adapted with GRNETs data and used for the participation to the ACROSS survey. Stakeholders receiving an invitation email to participate to the ACROSS survey and before receiving the link to the questionnaire, were asked to reply that they agree to the processing of their personal data as explained in the attached text, by sending a reply with the text: "I agree". For in-person sessions, the participants initially had to sign the consent form and then the session started.

a. Citizens Survey:

For the 45 citizens who agreed to the previous consent form, two different approaches were followed by GRNET. The first approach was asynchronous and online. The citizens were receiving by email a link to the respective questionnaire and a password to access it anonymously. The second approach was to organize in-presence sessions with teams of citizens (e.g., students) in a lab, where they filled in the questionnaire after a short introduction of what the ACROSS solution is.

b. Experts Survey:

The approach used for the 13 experts (policy makers, service providers, IT experts, business/NGO officials etc.) was as follows. A live or online session was organized with small teams of experts where ACROSS team members, from the respective Partner, showed them the prepared presentation and discussed with them about the general idea of ACROSS and the future of the service delivery and mobility in Europe. Afterwards a link to their respective questionnaire along with a password was sent to them to fill it in, anonymously, later.



2.1.2 Germany

1. Use Case Scenarios & Roadmap:

The aim of the user research process for the German use case was – 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 could include actions, touch points, emotions, pain points, and phases. Even though, various forms of interaction were possible, Germany decided to focus on qualitative research, primarily one-on-one interviews. Informed consent was asked from the participants prior to their participation in the interviews. At least five one-on-one interviews per ‘user profile’ (work-related move vs. study-related move) were done. This minimum of ten conducted interviews were then 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). Interviewed people who moved both to and away from their own pilot country.

The conduction of the interviews mostly consisted of open questions. Each interview participant could answer freely. No one was steered in a particular direction. No one was influenced in any way to achieve a particular result that the interview partner would not want to talk about. No questions were asked which would steer the participant into revealing sensitive information (i.e. information concerning health, race, sexual interest, etc.).

In addition, the results were anonymous and labeled. Names or other personal characteristics were not reused. Notes were taken, but no transcription or recording was made.

The described user profiles for work & study abroad were also used in the following evaluation and impact assessment activities.

2. Use Case Evaluation and Impact Assessment – Initial:

The second user research phase took place in June 2022 with the objective of evaluating the usability of the ACROSS Alpha Version. For this purpose, a usability test was devised, comprising three components:

- a. A structured interview
- b. Usability testing with tasks to assess usability and user experience.
- c. A final survey to capture user feedback after completing the tasks in the usability test.



Seven German citizens participated in this usability test. Data from the participants were pseudonymously collected, and no personally identifiable information was gathered.

3. Use Case Evaluation and Impact Assessment - Final:

The final research phase was done with another usability test and an extensive survey. In a span of around three months, from December 2023 to February 2024, surveys in Germany were conducted through online and in-person meetings, as well as flexible access to an online survey. Various channels were employed to contact potential participants to meet the predetermined number of participants agreed upon by the pilots. The study was conducted with citizens and experts to evaluate the end user perspective as well as the perception of experts - like public & private service providers and public administration representatives.

a. Citizen Survey:

To ensure a diverse representation of citizens with experience studying or working abroad, a multichannel approach was adopted. Initial evaluation sessions were conducted via online meetings using internal channels within Dataport to contact potential participants, particularly through working student and trainee channels. Subsequent calls for participation were made to recruit more participants, aiming to reach the target of 40 citizens. The second call allowed participation via an online link, offering greater flexibility to attract more interested participants. Despite not offering compensation, recruitment proved challenging as the Citizen Surveys required approximately one hour of participants' time. Out of 1631 citizens contacted, 39 citizen surveys were completed. Informed consents were received from all participating citizens.

b. Expert_Survey:

The expert evaluation was done in two phases. During the initial testing phase, data collection was done with two methods. Firstly, an in-person event was held on December 5, 2023, involving approximately 15 Dataport experts, among whom 5 volunteered as testers. Secondly, 42 Dataport experts were solicited to participate in an expert survey via email, facilitated by an online link from December 13 to December 22, 2024, with 12 agreeing to participate. In the subsequent testing phase, external experts from ministries and universities were recruited to contribute to the expert survey through email invitations with an online link. Initially sent between January 8 and January 19, 2024, the deadline was extended to February 16, 2024, due to unmet targets within the initial timeframe.



In total, 111 potential participants were contacted throughout the duration of the study. Among these 12 completed survey submissions with informed consent.

2.1.3 Latvia

1. Use Case Scenarios & Roadmap:

Similarly to previously described use cases of Greece and Germany, Latvia collected 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 was collected. Informed consent was asked from the participants prior to their participation in the interviews. Personal information (i.e. the responses that can be attributed to individual participants) is not disclosed to parties outside of the consortium or made public and is managed according to the 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.) was collected from other national regulatory authorities. This includes 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 are 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 were collected from national regulatory authorities.

On 9 September 2022 VARAM organised internal co-creation session within ACROSS team (four people). Its purpose was to assess current functionalities of the platform and suggest a list of requirements mainly from the perspective of public institutions – service providers. Team was briefed about overall progress, previous user requirements, gap analysis, ACROSS platform presentation and further roadmap of the project. Based on that, participants were asked to spot problems and opportunities for current ACROSS platform development. The team concluded that there must be some more functionalities added to ensure better control of data and choices for users, as well as provide more trust through verification of user journeys. Users would benefit if they could have a preview of proposed user journey and assess the need or possibility of its execution before actual initiation.

2. Use Case Evaluation and Impact Assessment – Initial:

VARAM based definition of use case scenarios on qualitative research provided by outsourced interviews. This approach ensured full anonymity of findings and provided a diverse and balanced group of



interviewees. In total, 10 interviews were conducted where 5 interviewees represented Latvians going abroad and 5 – other EU citizens moving to Latvia. Based on the information received and extensive desk research, two scenarios of working in Latvia and studying in Latvia were modelled as process flows, and information was valuable to create common scenarios with typical services used across all three use cases.

VARAM teamed up with Riga Stradiņš University to reach out to potential ACROSS end users (students from EU/EEA member states) and performed the usability test on 13 October 2022. The usability test was devised by pilot partners and Lisbon Council, as following comprise components.

- c. A structured interview
- d. Usability testing with tasks to assess usability and user experience.
- e. A final survey to capture user feedback after completing the tasks in the usability test.

The event gathered 16 foreign students in Latvia (*13 from Germany, 1 from United Kingdom, 1 from Austria, and 1 from Luxembourg*). They were introduced to ACROSS project and work done so far as presented by VARAM. End users accessed the ACROSS platform simultaneously and performed a task as laid out in the usability test. Before that they were introduced with the Consent forms and were asked to sign. The signed consent forms were collected and copied and collected in the project folder.

3. Use Case Evaluation and Impact Assessment - Final:

The final research phase was done with different usability Test and Survey. Starting from December 2023 to February 2024, surveys were conducted online and in person meetings. Some with citizens - students, some with stakeholders - experts. Different channels were used to reach the number of potential participants.

The study was conducted to evaluate the end user perspective as well as the perception of experts - like public administration representatives and public & private service providers.

a. Citizens Survey:

In the first working group of citizens – students organized on 14th November 2023, 13 students and citizens participated onsite, mostly from non-European countries - Azerbaijan, India, Ukraine, etc., also some from Latvia. Students mostly studying in Latvia University of Faculty of Business and Management (LU).

The other group of 17 students and citizens was organized onsite on 5th December 2023. They are also studying in the LU within the Erasmus+ program and they were from different European countries, mostly from Spain, Italy and other European countries. Instead of 30 students 10



citizens from Latvia was involved in the survey. In total 40 citizens were involved in the ACROSS survey. Thanks to the Latvian University for supporting the project by contacting the students as well as the staff. All the consent forms were signed by participants by hand and copies made and placed in the project folder.

b. Experts Survey:

The last evaluation survey was organized online for experts. According to the compiled list, more than 20 e-mails were sent with general information and the test, with a request to complete the test. It was a great challenge to get stakeholders. As a result, 10 responses were received, mostly from representatives of the state administration institutions and one from an NGO. Also, the Consent forms were sent to the experts online and all of them were signed with the e-signature and collected in the project folder.

There were two situations when one person sent the Consent form first and did not fulfill Survey and another expert also sent the signed Consent form and have not finished (submitted the Survey).

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 contains manuscripts reporting the conducted scientific work in ACROSS which have been accepted for publication in peer-reviewed journals and conferences. Such data contain aggregate and statistical data, and describe the personas created in ACROSS, but no personal data from the co-creation sessions are included. All these publications 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 are stored in PDF format. Each document is 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 is 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) is applied for ensuring open access to these publications. According to this archiving policy the author(s) of the publication 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 is also 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 are stored in PDF format. For each deliverable it is provided: (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 can 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 is 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 is maintained on the relevant webpage of the project website. This webpage grants open access to the PDF file of each listed public deliverable once it is 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 is 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 is 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.

The partners provided 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.

To facilitate the search, partners also include keywords or key phrases describing the subject or content of the data, including relevant industry terms.

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

3.2 Making data openly accessible

Materials generated under the ACROSS project were disseminated in accordance with the Consortium Agreement. The project deliverables that were marked as 'PU' (public) in the Description of Action were 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 fell outside the scope of the open access strategy. These included 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, fell outside of the scope of the open access strategy.

Therefore, any data in PU deliverables was anonymized. This implies that co-creation responses as a data set (including any audio and video recordings) were entirely out of scope; and that survey outcomes were only reported on at the aggregate, statistical level. Such data is 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 is not 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 were shared only in a pseudonymized or anonymized manner. Original (non-anonymized or non-pseudonymized data) are stored in order to allow identification and traceability for research validations and follow-up, but such storage is organized 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 are restrictions on what data can be published openly. The primary way of accessing the data is via HTTP REST API. The APIs are specified



using the OpenAPI standard. The data and metadata are stored in their respective databases. The code is available in the official ACROSS Gitlab.

3.3 Making data interoperable

The metadata structure coheres to the DCAT-AP specification. Data is 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 is included in WP4 deliverables.

3.4 Increase data re-use

Specific details on future use of the data are 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, standardized and widely known license models, such as e.g. a 'CC-BY-SA 3.0' license, as a rule to all research data to facilitate the widest re-use as possible, or e.g. the **EUPL for software releases**.

All open research data are made available for re-use without any data embargo, meaning that all data are 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 were decided upon once the selection of data sources were finalized and thoroughly evaluated.



4 Allocation of resources

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

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

The GitLab public repository can be accessed through the following link [public · GitLab \(tecnalia.com\)](#)

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 adheres to security – by – design and privacy – by design principles that allow the security and audit standards to remain consistent across multiple environments.

Furthermore, ACROSS implemented 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 personally identifiable information.

More information regarding the security measures implemented by each Consortium partner can be found in D8.8 (which was updated later in the project based on comments received by the reviewers in the first review).

The architecture is depicted 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:

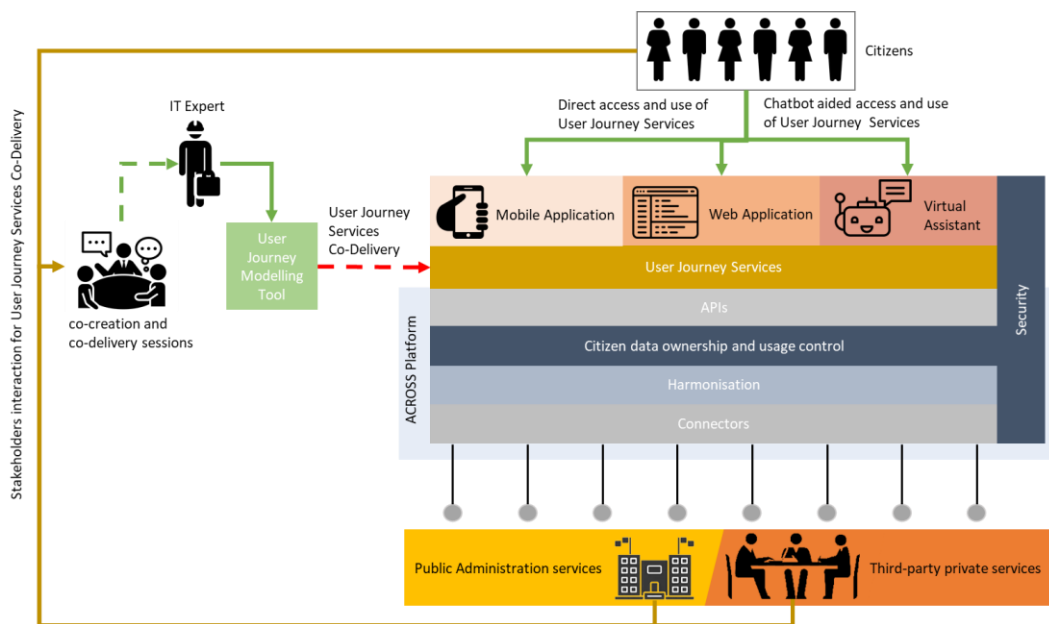


Figure 1 - ACROSS high-level architecture (Source: DOA)

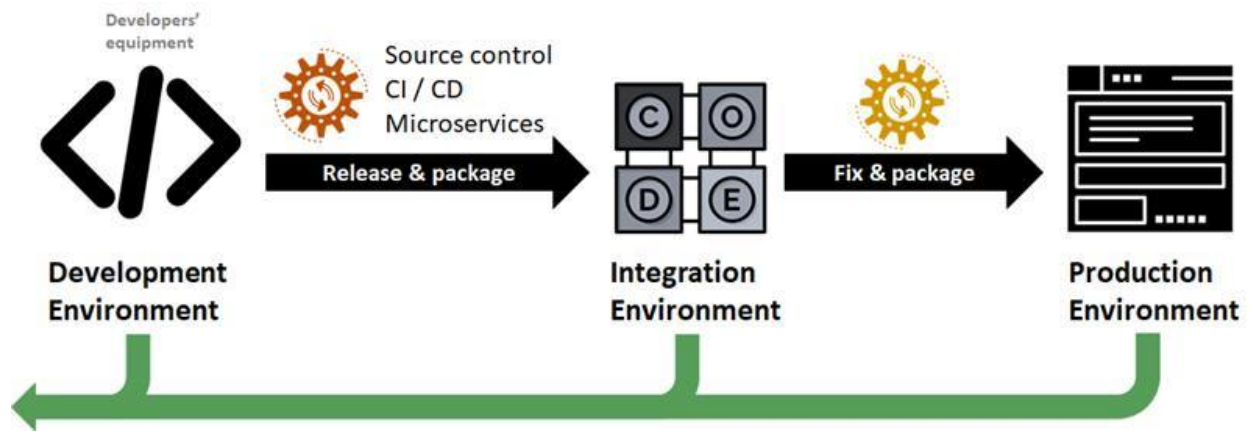


Figure 2 - The environments in ACROSS

At the development stage each partner managed the security of their development environments. The source code was uploaded to GitLab, hosted in TECNALIA, which follows the LDAP [7] authentication protocol. The integration environment includes 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 need to be installed in the use cases, just those that are of interest for the pilot cities. The security in this case follows the policies defined in the different municipalities.



6 Ethical Aspects

The ethical aspects of the ACROSS project were assessed under Work Package 8, which set 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. After having received comments from the reviewers during the first review, the Consortium has appointed an external and independent ethics advisor for the ACROSS project.
- 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 for the Consortium in general and specifically for the partners that are involved in data collection.
- The definition of anonymization and pseudonymization strategy.
- Explicit confirmation that the beneficiary has lawful basis for the data processing, and that the appropriate technical and organizational 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 Conclusions

The deliverable at hand has presented the plan for the data management in the ACROSS project. In this action, different types of data were 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 is either anonymous, and from open data repositories, or sought to be anonymized as much as possible. Whenever personal data is collected, explicit consent was required, and data was thereafter anonymized as soon as reasonably practicable.

All data and metadata generated followed machine-readable formats.

Data from the use cases was stored on the use cases' premises fulfilling the relevant legislation. Data from publications is stored in OpenAire indexed repositories favoring the green model whenever possible. Other publications such as deliverables are stored at TECNALIA's hosting services.



8 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 . [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
- [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/>.
- [8] ACROSS Consortium, D6.1 Use Cases scenarios and Roadmap
- [9] ACROSS Consortium, D6.2 Use case evaluation and impact assessment (initial)
- [10] ACROSS Consortium, D6.3 Use case evaluation and impact assessment (final)