RAINBOW

| | |
|---|---|
| Project Title | AN OPEN, TRUSTED FOG COMPUTING PLATFORM FACILITATING THE DEPLOYMENT, ORCHESTRATION AND MANAGEMENT OF SCALABLE, HETEROGENEOUS AND SECURE IOT SERVICES AND CROSS-CLOUD APPS |
| Project Acronym | RAINBOW |
| Grant Agreement No | 871403 |
| Instrument | Research and Innovation action |
| Call / Topic | H2020-ICT-2019-2020 /<br>Cloud Computing |
| Start Date of Project | 01/01/2020 |
| Duration of Project | 36 months |

# D7.11 – Open Research Data Pilot Contribution

| | |
|---|---|
| Work Package | WP7 – Dissemination, Exploitation and Communication |
| Lead Author (Org) | Georgios Kakamoukas (K3Y) |
| Contributing Author(s) (Org) | Christina Stratigaki (UBITECH) |
| Due Date | 31.03.2020 |
| Actual Date of Submission | 31.03.2020 |
| Version | V1.0 |

**Dissemination Level**

| | |
|---|---|
| X | PU: Public (*on-line platform) |
| | PP: Restricted to other programme participants (including the Commission) |
| | RE: Restricted to a group specified by the consortium (including the Commission) |
| | CO: Confidential, only for members of the consortium (including the Commission) |

European Commission

The work described in this document has been conducted within the project RAINBOW. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 871403. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

## Versioning and contribution history

| Version | Date | Author | Notes |
|---|---|---|---|
| 0.1 | 18.03.2020 | Georgios Kakamoukas (K3Y) | First Draft |
| 0.6 | 26.03.2020 | Georgios Kakamoukas (K3Y) | Second Draft |
| 0.9 | 27.03.2020 | Christina Stratigaki (UBITECH) | Third draft, Section 2 |
| 0.95 | 30.03.2020 | George Pallis (UCY) | Peer Review |
| 1.0 | 31.03.2020 | Christina Stratigaki (UBITECH), George Kakamoukas (K3Y) | Formatting and Final version |

**Disclaimer**

This document contains material and information that is proprietary and confidential to the RAINBOW Consortium and may not be copied, reproduced or modified in whole or in part for any purpose without the prior written consent of the RAINBOW Consortium

Despite the material and information contained in this document is considered to be precise and accurate, neither the Project Coordinator, nor any partner of the RAINBOW Consortium nor any individual acting on behalf of any of the partners of the RAINBOW Consortium make any warranty or representation whatsoever, express or implied, with respect to the use of the material, information, method or process disclosed in this document, including merchantability and fitness for a particular purpose or that such use does not infringe or interfere with privately owned rights.

In addition, neither the Project Coordinator, nor any partner of the RAINBOW Consortium nor any individual acting on behalf of any of the partners of the RAINBOW Consortium shall be liable for any direct, indirect or consequential loss, damage, claim or expense arising out of or in connection with any information, material, advice, inaccuracy or omission contained in this document.

## Table of Contents

## List of tables

## List of figures

# Executive Summary

RAINBOW project participates in Open Research Data Pilot carried out by the European Commission, as it is willing to support the principles of Open Access to research data and publications generated through H2020 programmes. Consequently, in D7.1 an overview of Data Management Plan has been created to deal with the data collected and generated during the project. Current deliverable aims to present the tools that will allow RAINBOW to contribute to ORDP and details the Data Management procedures.

# 1 Introduction

The Open Research Data Pilot (ORDP) of the European Commission enables open access and reuse of research data generated by Horizon 2020 projects. There are two main pillars to the Pilot: a) developing a Data Management Plan (DMP) and b) providing open access to research data.

A project that opts-in ORDP have to adhere to the following conditions:

- Develop (and keep up-to-date) DMP.
- Deposit the data in a research data repository.
- Ensure third parties can freely access, mine, exploit, reproduce and disseminate this data.
- Provide related information and identify (or provide) the tools needed to use the raw data to validate the research.

The ORDP applies to:

- The data (and metadata) needed to validate results in scientific publications.
- Other curated and/or raw data (and metadata) that are specified in the DMP

Considering privacy and data protection issues scientific research data should be easily discoverable, accessible, assessable and intelligible, useable beyond the original purpose for which it was collected and interoperable to specific quality standards.

Project piloting in the Open Research Data activity, has to consider the following aspects: Regarding the digital research data generated in the action ('data'), the beneficiaries must deposit those in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

- the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
- other data, including associated metadata, as specified and within the deadlines laid down in the data management plan.

The RAINBOW Data Management also follows the Guidelines on FAIR Data Management in Horizon 2020, released by the European Commission Directorate – General for Research & Innovation. This Horizon 2020 FAIR DMP template[1] has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. According to these guidelines the management and organization of data should be based on four basic principles, which determine how research outputs should be processed so that they can be more easily accessed, understood, exchanged and reused. This means that data must be findable, accessible, interoperable and re-useable, for example by researchers interested in using the data in further research in the field.

These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution. EC provides a Template with the FAIR principle. This template is not intended as a strict technical implementation of

[1] H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, Version 3, 26 July 2016

the FAIR principles, it is rather inspired by FAIR as a general concept. The template represents the set of questions that someone should answer with a level of detail appropriate to the project.

The DMP creation will be based on a cooperative activity where each partner will examine whether they foresee to produce data to which open access can be granted.

## 1.1 Definitions

**Open Access**: Open access means unrestricted access to research results. Often the term open access is used for naming free online access to peer-reviewed publications. Open access is expected to enable others to:

a) build on top of existing research results;
b) avoid redundancy;
c) participate in open innovation; and
d) read about the results of a project or inform citizens.

All major publishers in computer science - like ACM, IEEE, Elsevier, or Springer - participate in the idea of open access. Both green, or gold open access levels are promoted. Green open access means that authors eventually are going to publish their accepted, peer-reviewed articles themselves, e.g. by deposing it to their own institutional repositories. Gold open access means that a publisher is paid (e.g. by the authors) to provide immediate access on the publishers' website and without charging any further fees to the readers.

**Open Research Data**: Open research data is related to the long-term deposit of underlying or linked research data needed to validate the results presented in publications. Following the idea of open access, all open research data needs to be openly available, usually meaning online availability. In addition, standardized data formats and metadata has to be used to store and structure the data. Open research data is expected to enable others to:

a) understand and reconstruct scientific conclusions; and
b) to build on top of existing research data.

**Metadata**: Metadata defines information about the features of other data. Usually metadata is used to structure larger sets of data in a descriptive way. Typical metadata are names, locations, dates, storage data type, and relations to other data sets. Metadata is very important when it comes to index and search larger data sets for a specific kind of information. Sometimes metadata can be retrieved automatically from a dataset, but often it needs some manual classification also. The well-known tags in MP3-recordings are a good example why metadata is necessary to find a specific kind of genre or composer in a larger number of songs.

# 2 FAIR data

RAINBOW project supports the reuse of research data and follows FAIR principles[2]. FAIR represents a set of guiding principles to make data **Findable, Accessible, Interoperable, and Reusable**.

The international FAIR Principles have been formulated as a set of guidelines for the reuse of research data. The acronym FAIR stands for findable, accessible, interoperable and reusable. Research data must be of quality that makes them accessible, findable and reusable.

- **Findable:** data has a unique, persistent ID, located in a searchable resource, and documented with meaningful metadata.
- **Accessible:** data is readily and freely retrievable using common methods and protocols, metadata is accessible even if the data is not.
- **Interoperable:** data is presented in broadly recognized standard formats, vocabularies, and languages.
- **Re-useable:** data has clear licenses, and accurate meaningful metadata conformity to relevant community standards and identifying its content and provenance.

## 2.1 Making data findable, including provisions for metadata

This document launches the data management plan to support the effective collection and integration of the RAINBOW data. Storage, processing and sharing (among project participants) will occur via data exchange platforms (such as Nextcloud), whereas interaction with the wider public will be achieved through the official project website. Also, data will be stored at the coordinator's repository and will be kept for minimum 5 years after the end of the project. Where requested, data will be kept for 2 more years.
A naming convention will include a concise description of contents, the host institution collecting the data and the month of publication.

Version numbering will only be an issue if a participant requests withdrawal of their data in which case a version number will be added to the filename.

No specific standards or metadata have been identified for the time being for the proposed datasets.

Data will be anonymized meaning that data will not identify any individuals and therefore real names of participants will NOT be distributed.

Data will be shared only in relation to publications (deliverables and papers). As such, the publication will serve as the main piece of metadata for the shared data. When this is

[2] Force11 (2016) The FAIR Data Principles, https://www.force11.org/group/fairgroup/fairprinciples

not seen as being adequate for the comprehension of the raw data, a report will be shared along with the data explaining their meaning and methods of acquisition.

### 2.1.1 Discoverability of the data

In order to be able to use the data generated by the project is essential to integrate data from the participants in the open calls and the activities undertaken by project partners. Taking into account the FAIR data principles (Wilkinson et al., 2016[3]) (meta)data should:

- Be assigned to a globally unique and persistent identifier;
- contain enough metadata to fully interpret the data, and;
- be indexed in a searchable source.

By applying these principles data become retrievable and include their authentication and authorization details.

### 2.1.2 Data identification mechanisms

All documents associated project will be identified with a project name and unique and persistent document type designator and number that will be given to the coordinator for the submission to the EC. Versioning of the document should be part of the document name and title.

As per the documents related to project activities and/or deliverables, the tasks or deliverables number will be used to identify the document followed by a brief title of the activity or deliverable.
Examples:

- RAINBOW-D10.1-Project Web Portal-v1.0.pdf
- RAINBOW-D2.4-Data Management Plan-v1.0.pdf

### 2.1.3 Naming conventions used

Each set of data produced (dataset, deliverables, etc.) will be named in a uniform way and will include a table with a version control.

The recommendations to name documents of the project are as follows[4]:

- Choose easily readable identifier names (short and meaningful);
- Do not use acronyms that are not widely accepted;
- Do not use abbreviations or contractions;
- Avoid Language-specific or non-alphanumeric characters;
- Add a two-digit numeric suffix to identify new versions of one document.
- Dates should be included back to front and include the four-digit years: YYYYMMDD.

[3] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data3:160018 doi: 10.1038/sdata.2016.18 (2016).
[4] https://www.ukdataservice.ac.uk/manage-data/format/organising

For deliverables: **RAINBOW_[Deliverable Code]-[Deliverable Title]_[Partner]-vA.BB**
i.e.: RAINBOW_D2.1-Project Management Handbook_UBITECH-v1.00 *(for submission to the Commission)*

For datasets: **WP [Work Package number] P [Pilot number; pilot activity number] - [description of the activity]** i.e.: WP6 P1.3 Results of demonstration performance.

The keywords used to easily identify documents related to a project will be the ones used throughout the submission process, where applicant will have to select the characteristics of their projects selecting descriptors from a dropdown menu.

### 2.1.4 Clear versioning of the documents

Only documents created by the consortium will be versioned, for this purpose templates include 3 descriptors to identify the versions and status of the documents:

*Table 2-1 – Proposed Document History Table overview*

| Version | Date | Comments | Author |
|---|---|---|---|
| 1 | xx | xx | xx |
| 2 | | | |
| 3 | | | |
| 4 | | | |

Moreover, partners, following the recommendations included in section "Naming conventions" will identify the different versions by using a two-digit number following the descriptor Draft. A document reviewed by another partner should be returned to the principal author by including **rev + acronym** of the organisation. Only the principal author will change the draft number and will add the word FINAL to documents ready to be sent to the EC or those to be used as final versions.
The document history included in the document template should be filled in as follows:

| Version | Date | Comments | Author |
|---|---|---|---|
| 1 | XX/XX/2020 | Section 2.1 needs to be completed | ABC |
| 2 | XX/XX/2020 | Section 2,1 completed. Comments added to the document. | CDE |
| 3 | XX/XX/2020 | Added suggestions by EMAX | ABC |
| 4 | XX/XX/2020 | Included some topics on section 2.1 | XYZ |
| | XX/XX/2020 | Final version with partners contribution | ABC |

### 2.1.5 Standards for metadata creation (if any)

Basic metadata will be used to facilitate the efficient recall and retrieval of information by project partners and external evaluators and contribute to easily find the information requested. To this end, all documents related to the project have to include in the front-page information about author(s) & editor(s), WP, dissemination level and version.

## 2.2 Making data openly accessible

Where possible data will be made available subject to Ethics and participant agreement. However, the personally-identifiable nature of the data collected within RAINBOW means that in most instances it would be difficult to release collected data. Where data is made available, we will do so using the coordinator's repository.

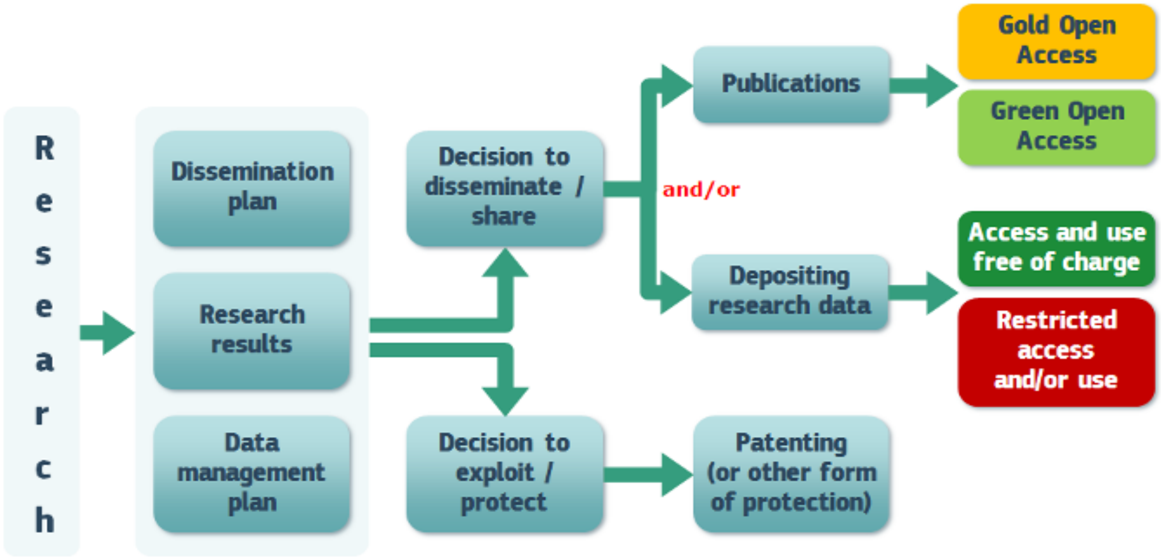
Research
Dissemination plan
Research results
Data management plan
Decision to disseminate / share
Decision to exploit / protect
and/or
Publications
Depositing research data
Patenting (or other form of protection)
Gold Open Access
Green Open Access
Access and use free of charge
Restricted access and/or use

*Figure 2-1 Open access to scientific publication and research data in the wider context of dissemination and exploitation*[5]

Prior to release, a requesting party will need to contact the Project Coordinator describing their intended use of a dataset. The Project Coordinator will send a terms and conditions document for them to sign and return. Upon return, the dataset will be released. Documentation will be included with the release of the data.

### 2.2.1 Methods or software needed to access the data

No specific software tools will be needed to access the data, since anonymous data sets will be saved and stored in word, pdf or excel to facilitate its exploitation and guarantee their long-term accessibility.

### 2.2.2 Deposit of data, associated metadata, documentation and code

Data will be deposited and secured on Nextcloud platform and additional instance of all data on coordinator's account.

[5] European Commission Directorate-General for Research & Innovation (2017) Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

## 2.3 Making data interoperable

The concept interoperable demands that both data and metadata must be machine-readable and that a consistent terminology is used.

### 2.3.1 Interoperability of data assessment

Partners will be responsible of storing the data in a comprehensive format and adapted to the real and current needs of the possible practitioners interested in using, merging or exploiting the data generated throughout the project. The assessment of data interoperability will be updated in future reviews in order to guarantee the RAINBOW data fits the needs of a specific scenario (such as data infrastructures, interests or purpose of data.

### 2.3.2 Vocabulary use

The vocabulary used in the project is a very standard and common language within the business creation culture and the logistics. Vocabulary won't represent any barrier for data interoperability a re-use.

### 2.3.3 Increase data re-use through clarifying licenses

Due to the sensitive nature of the data they will only be available on application/Nextcloud platform/share portal and their use will be restricted to the research use of the licensee and colleagues on a need-to-know basis. This non-commercial licence is renewable after 2 years, data may not be copied or distributed and must be referenced if used in publications. These arrangements will be formalised in a User Access Management licence which describes in detail the permitted use of the data.

### 2.3.4 Data quality assurance process

The project coordinator will be responsible of assuring the quality of the data by making sure dataset follow the FAIR principles included in this plan, and that data is updated.
Personal data processing will be done following the EU, national and international laws taking into account the "data quality" principles listed below[6]:
Data processing is adequate, relevant and non-excessive;

- Accurate and kept up to date;
- Processed fairly and lawfully;
- Processed in line with data subjects' rights;
- Processed in a secure manner;
- Kept for no longer that necessary and for the sole purpose of the project.

Data quality assurance process will be led in accordance with the REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

[6] Wilms, G. Guide on Good Data Protection Practice in Research of the European University Institute. (March 2017). Retrieved from http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf

### 2.3.5 Length of time for which the data will remain re-usable

The Consortium will contribute to maintain data re-usable as longer as possible after the end of the project. A first period of 4 years has been established; however, this time can be extended under partners agreement. This period can vary depending on the value of the data after the end of the project.

# 3 Data collection

The following table summarizes the procedures for collecting project related data.

*Table 3-1 Data Usage Scenarios*

| Data usage scenarios | Nature of Datasets | | |
|---|---|---|---|
| | Confidential | Anonymized and Public | Non anonymized |
| Original data produced by the consortium | Surveys<br>Interviews<br>Pilot activities: workshops, meeting with stakeholders, co-design sessions; evaluation sessions;<br>F2F / distant interaction | Newsletters<br>Publications<br>Personal Emails<br>Open Access repositories | Events coverage – directly or via specialised agencies<br><br>A/V conferencing systems<br>Internal repositories |
| Existing data already in possession of the consortium and/or partners | Seamless access and use during project execution | Seamless access and use during project execution | N/A |
| Existing data sourced/procured by the consortium and/or partners | Licensed access and use during project execution | Free and open access and use during project execution | N/A |

Every partner is responsible for the behaviour of all team members, which may also include subcontracted organizations (e.g. specialized press agencies) or even volunteers. The latter circumstance does not exempt the delegate of a certain job in case of improper application of extant norms and rules.
All data will be collected in a digital form – therefore CSV, PDF, Word, xls spreadsheets and textual documents will be the prevalent formats. In case of audio/video recordings and images, the most appropriate standards will be chosen and adopted (such as .gif, .jpg, .png, .mp3, .mp4, .mov and .flv). Website pages can be created in .html and/or .xml formats.
All data collection and processing tasks within RAINBOW consortium will be GDPR compliant and will be carried out according to EU and national legislation. RAINBOW partner should inform RAINBOW Ethics Committee on the planned data collection and/or processing and describe its purpose. Ethics Committee will review the request

and will give feedback to the partner so as to follow the due process. Any data collected by RAINBOW consortium will be anonymised/pseudonymised and the partner must explain how all the data they intend to process is relevant and limited to the purposes of their intended task (in accordance with the 'data minimisation' principle). Ethics Committee will review the request and will give feedback to the partner so as to follow the due process. Also, the Ethics Committee will evaluate the ethics risks related to the data processing activities of the project, including conducting data protection impact assessments under art.35 GDPR. Any confidential data collected from the users will be handled only by the involved partners using local data management and storage systems, obtainable with different levels of access, regulated by the Partner acting as Controller of the data. The Partner acting as Controller will also be responsible for data management, secure storage and deletion of the confidential data beyond the lifetime of the project. The publicly available data  will be stored on GDPR compliant Cloud platform, with different levels of access, regulated by the Project coordinator. Any data that is publicly available will be accessible through FAIR (Findability Accessibility Interoperability and Reusability) principle. RAINBOW Consortium will submit an explicit confirmation to Ethics Committee that the data is publicly available and can be freely used for the purposes of the project. At the completion of the project, all the responsibilities concerning long-term data management and secure storage of the publicly available data will fall on the selected service for storing project data based on the decision and agreement established by the consortium.

## 3.1 Human-Robot Collaboration in Industrial Ecosystems Use Case Data Procedures

The use case 'Human-Robot Collaboration in Industrial Ecosystems' enables several innovative location-based services, because such accuracy levels essentially allow for real-time interaction between humans and cyberphysical systems. Activity recognition, machine navigation (e.g., “shelf” level), geo-fencing, and automated robotics; are among services that yield safety-critical assembly processes and logistics.

RAINBOW will deploy BIBA indoor positioning services to physical fog nodes that span across manufacturing factories with the task of processing, structuring and normalizing sensing data and then performing in place analysis to derive the coordination plan and collision detection. Combining the aforementioned information, fog nodes can prevent collisions and fatal accidents. Fog nodes are placed near assembly lines, factory sections or floors, and may be mobile, moving across the factory with the sole purpose of RAINBOW guarantying deterministic and in time reaction. By adopting fog nodes, processing is distributed and independent per decomposed unit (e.g., assembly line), with fog nodes forming an overlay mesh network to share data analysis results and aggregate data for further assessment at factory and area level.

This use case will not get access or produce any personal data.

## 3.2 Digital Transformation of Urban Mobility Use Case Data Procedures

As stated in the Grant Agreement in page 154, RAINBOW will take advantage of the open data that are available by the “smart city infrastructure of the city of Torino which includes: smart traffic lights, road sensors, air and noise quality, weather conditions (https://www.torinocitylab.it/en/). AperTo, the city's open data portal, is a general repository that shares internal data of the administration following the "open data" rules, respecting the directive 2013/37/UE (PSI2).

AperTo uses the CKAN platform, a content management system (CMS) that follows the DCAT-AP IT metadata standard of AgiD. AperTo supports CKAN's API for both data exposition and harvesting in "machine to machine" mode and grants a system with a visual and multiuser approach to the data publication. The data published on the platform are gathered and released by different city offices who have responsibility upon them. All the data can be released through Creative Commons Attribuzione 4.0 (CC BY) licensing. Firms who test through Torino City Lab have access to the portal and to the township's open data team. Whenever a firm uses data from the portal, they can then post the results of their work on the portal itself. The main goal is to give visibility to the work of every innovator and enlarge the dataset that can be used by our community.

The reporting of geo-referenced data from vehicles and other sources (e.g., crowdsourced citizens' reports) will be done through the use of advanced Direct Anonymous Attestation (DAA) mechanisms in order to test in parallel the use of DAA for enhanced privacy provision; all reports will be anonymously signed from the devices' TPM (secret keys will be created and securely stored in the TPM), thus, allowing us to achieve user-controlled anonymity, unlinkability, nontraceability, unforgeability and non-frameability.

## 3.3 Power Line Surveillance via Swarm of Drones Use Case Data procedures

### 3.3.1 Privacy Protection Risks and Their Countermeasures

Use case assumes the use of drones to collect images of power lines in order to assess their technical condition. As with every image collection in an uncontrolled environment there is a chance that images will contain not only the main subject (power lines and pylons), but also pedestrians and vehicles. This brings a risk that the system might capture and process images that contain bystander faces or license plates of their vehicles. In order to mitigate this risk, we propose three countermeasures that will be employed sequentially:

1. areas for test flights will be chosen in such a way that the probability of encountering a pedestrian or third-party vehicle will be reduced to minimum,
2. drone flight level and camera parameters will be chosen in such way that it will be possible to capture images of power lines and power pylons that will be usable for technical condition assessment, yet objects in the background will not be photographed with the same level of detail, thus will not recognizable,
3. captured images will be automatically anonymized before they will be stored in the cloud.

### 3.3.2 Test flight area selection

Areas in which test flights will be performed must be carefully selected not only because of privacy concerns, but first of all because of safety. The best candidates are rural areas, where power lines are crossing fields or forests, away from roads and buildings. This reduces the risk of collision with a human or causing damage to property in case of a drone failure. The criteria for choosing the test flight area from the safety point of view, also reduce the privacy violation risk.

MSP experience with power line inspection data collection shows that it is quite easy to select such areas that the probability that a person or third-party vehicle will be accidentally photographed is almost zero. The most probable causes that this would happen are that there are farmers working in the field or linemen working on the power infrastructure. In such an event one can simply relocate to an alternative test flight area.

Below we show exemplary footage captured with a drone of similar class to the one planned to be used during test flights in the use case. The three areas in which those images were captured meet the criteria that minimize the risk of privacy violation.

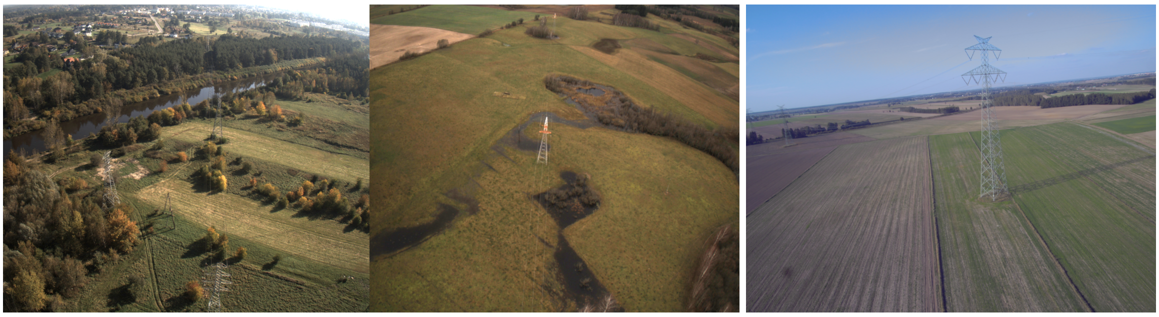

*Figure 3-1 Test images*

First two images were captured on much higher altitude than planned in the use case, therefore the camera had a bigger field of view. With lower altitude, it is possible to select such a section of the power line where there are no houses nor roads (and thus potential bystanders or third-party vehicles) are visible.

### 3.3.3 Flight level and camera parameters

Even after careful selection of the test flight area, there is a slight chance that still some person or third-party vehicle might stumble into the camera field of view. This can only be an issue if the resolution of the photo is high enough for that person (or vehicle's licence plate) to be recognizable. The spatial resolution in the image (size of a pixel in meters) varies with the distance from the camera. Power grid operators require that the photos captured for the power line inspection should have resolution of around 300 pixels per meter at the height of the power line. If we assume that the drone will fly 10 meters above the power line and that the typical height of a power pylon is 30 meters, from triangle similarity, it can be easily calculated that at the height of a human face the spatial resolution will be less than 80 pixels per meter. Simulation of a photo that would be produced is shown below (enlarged):

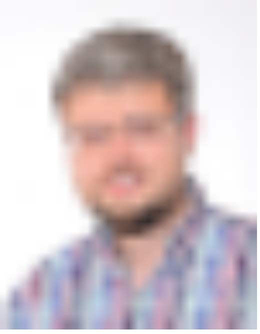

*Figure 3-2 photo resolution example*

Also, camera lenses have limited depth of field (if focused at a point closer than its hyperfocal distance). This means that only objects within a certain range of distances from the camera will be in focus and all closer or further objects will be blurred and thus unrecognizable.

During test flight planning, the flight level and focus setting of the camera lens will be adjusted in such a way that captured photos will be still usable for inspection, but the possibility to recognize people or licence plates on the ground will be minimized.

### 3.3.4 Automatic image anonymization

If for some reason, despite the counter measures described in the previous two points, the drone would capture an image of a person or third-party vehicle, a third technique will be employed - all images will be automatically anonymized before they will be stored. Both the camera and the drone do not store the captured images permanently, nor process them. For the processing the images are transferred to the Ground Control Station along with the information about the orientation and the position of the drone. Before the image is processed, displayed to the operator or transferred to the cloud for permanent storage, it will be automatically anonymized. This can be achieved by using CNN-based face and licence plate detectors that will automatically recognize the regions containing sensitive data and then these regions will be blurred. The drone orientation and position data itself does not pose privacy violation risk, so it can be immediately distributed via Rainbow network to other ground control stations.

## 3.4 Digital Transformation of Urban Mobility Data Procedures

As stated in the Grant Agreement in page 154, RAINBOW will take advantage of the open data that are available by the "smart city infrastructure of the city of Torino which includes: smart traffic lights, road sensors, air and noise quality, weather conditions (https://www.torinocitylab.it/en/). AperTo, the city's open data portal, is a general repository that shares internal data of the administration following the "open data" rules, respecting the directive 2013/37/UE (PSI2).

AperTo uses the CKAN platform, a content management system (CMS) that follows the DCAT-AP IT metadata standard of AgiD. AperTo supports CKAN's API for both data exposition and harvesting in "machine to machine" mode and grants a system with a visual and multiuser approach to the data publication. The data published on the platform are gathered and released by different city offices who have responsibility upon them. All the data can be released through Creative Commons Attribuzione 4.0 (CC BY) licensing.

Firms who test through Torino City Lab have access to the portal and to the township's open data team. Whenever a firm uses data from the portal, they can then post the results of their work on the portal itself. The main goal is to give visibility to the work of every innovator and enlarge the dataset that can be used by our community.

The reporting of geo-referenced data from vehicles and other sources (e.g., crowdsourced citizens' reports) will be done through the use of advanced Direct Anonymous Attestation (DAA) mechanisms in order to test in parallel the use of DAA for enhanced privacy provision; all reports will be anonymously signed from the devices' TPM (secret keys will be created and securely stored in the TPM), thus, allowing us to achieve user-controlled anonymity, unlinkability, nontraceability, unforgeability and non-frameability.

# 4 Publishing Infrastructure for Open Access

The RAINBOW publication infrastructure consists of a process and several web-based publication platforms that together provide long-term open access to all publishable, generated or collected results of the project. The implementation of the project fully complies with law in national and EU level and especially with the Directive 95/46 related to the Protection of personal data. More specifically, there are not cases where personal data information or sensitive information of internet users is collected (IP addresses, email addresses or other personal information) or processed. For the whole duration of the project, from the beginning to its end, the RAINBOW Ethics Committee will carefully examine the legality of the activities and the tools (including platforms) that will be produced for not violating the personal data of internet users. In the potential future case where the RAINBOW consortium will collect, record, store or process any personal information, it will be ensured that this will be done on a basis of respecting citizens' rights, preventing their identification and keeping their anonymization. Both the process and the used web-based platforms are described in the following subsections. The section ends with some research items and data that they project partners foresee that they will be reflected on respective publication actions. The research data will be the main subject of analysis with respect to data management in the next section.

## 4.1 Publishing Process

RAINBOW partners defined a simple, deterministic process that decides if a result in RAINBOW must be published or not. The term result is used for all kind of artefacts generated during RAINBOW like white papers, scientific publications, and anonymous usage data. By following this process, each result is either classified public or non-public. Public means that the result must be published under the open access policy. Non-public means that it must not be published.

For each result generated or collected during RAINBOW runtime, the following questions must be answered to classify it:

## 4.2 Publishing Process

***Does a result provide significant value to others or is it necessary to understand a scientific conclusion?***

If this question is answered with yes, then the result is classified as public. If this question is answered with no, the result is classified as non-public. Such a result could be code that is very specific to RAINBOW platform (e.g., a database initialization) which is usually of no scientific interest to anyone, nor does it add any significant contribution.

***Does a result include personal information that is not the author's name?***

If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published. This also bare witness on the repetitive nature of the publishing process, where results which are deemed in the beginning as non-publishable can become publishable once privacy-related information is removed from them.

***Does a result allow the identification of individuals even without the name?***

If this question is answered with yes, the result is classified as non-public. Sometimes data inference can be used to superimpose different user data and reveal indirectly a single user's identity. As such, in order to make a result publishable, the included information must be reduced to a level where single individuals cannot be identified. This can be performed by using established anonymization techniques to conceal a single user's identity, e.g., abstraction, dummy users, or non-intersecting features.

***Does a result include business or trade secrets of one or more partners of RAINBOW?***

If this question is answered with yes, the result is classified as non-public, except if the opposite is explicitly stated by the involved partners. Business or trade secrets need to be removed in accordance to all partners' requirements before it can be published.

***Does a result name technology that is part of an ongoing, project-related patent application?***

If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after patent has been filed.

***Can a result be abused for a purpose that is undesired by society in general or contradict with societal norms and RAINBOW's ethics?***

If this question is answered with yes, the result is classified as non-public.

***Does a result break national security interests for any project partner?***

If this question is answered with yes, the result is classified as non-public.

## 4.3 Publishing Platforms

In RAINBOW, we use several platforms to publish our results openly. The following list presents the platforms used during the project and describes their concepts for publishing, storage, and backup.

### 4.3.1 Project Website

The partners in the RAINBOW consortium decided early to setup its own project-related webpage, which has been set up since the third month of the project. Its purpose is to describe the mission and the general approach of the project and its development status, as well as provide a short description of the project's objective and its methodology, post news, events and updates that are relevant to the project's activities. A dedicated page for project's public documents is available where all the deliverables of the project are published in portable document format (PDF). The webpage was designed by the technical coordinator (K3Y) and was reviewed by the whole RAINBOW project consortium. All webpage-related data is backed on a regular basis. All information on the RAINBOW website can be accessed without creating an account. The webpage is backed manually every two weeks. The RAINBOW Project Portal will be collecting useful (anonymous) data regarding its visitors including unique visitors, countries of origin, time spent on portal etc., and will be available during the project runtime, and for at least two years after the official project end.
Website available at: https://rainbow-h2020.eu/

### 4.3.2 Zenodo

Zenodo is a research data archive / online repository which helps researchers to share research results in a wide variety of formats for all fields of science. It was created through EC's OpenAIRE+ project and is now hosted at CERN using one of Europe's most reliably hardware infrastructures. Data is backed nightly and replicated to different locations. Zenodo not only supports the publication of scientific papers or white papers, but also the publication of any structured research data (e.g., using XML). Zenodo provides a connector to GitHub that supports open collaboration for source code and versioning for all kinds of data. All uploaded results are structured by using metadata, like for example the contributors' names, keywords, date, location, kind of document, license, and others. Considering the language of textual metadata items, English is preferred. All metadata is licensed under CC0 license (Creative Commons 'No Rights Reserved'). The property rights or ownership of a result does not change by uploading it to Zenodo.
All public results generated or collected during the project lifetime will be uploaded to Zenodo for long-term storage and open access.
Web-Link: http://zenodo.org

### 4.3.3 Github

GitHub is a well-established online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables world-wide collaboration between developers and provides also some facilities to work on documentation and to track issues. GitHub provides paid and free service plans. Free service plans can have any number of public, open-access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan. Many open-source projects use GitHub to share their results for free. The platform uses metadata like contributors' nicknames, keywords, time, and data file types to structure

the projects and their results. The terms of service state that no intellectual property rights are claimed by the GitHub Inc. over provided material. For textual metadata items, English is preferred. The service is hosted by GitHub Inc. in the United States. GitHub uses a rented Rackspace hardware infrastructure where data is backed up continuously to different locations.
All source-code components that are implemented during this project and decided to be public will be uploaded to an open access GitHub repository.
Web-Link: https://github.com/

### 4.3.4 ResearchGate Channel

A RAINBOW ResearchGate channel will be established to promote the dissemination of scientific publications of the project. Open Access documents are published using the portable document format (PDF). All downloads are enriched by using simple metadata information like the title, a short description and the type of the document.
The RAINBOW ResearchGate channel is managed by K3Y, while all partners periodically update the material. The link for accessing the RAINBOW ResearchGate channel is: https://www.researchgate.net/project/RAINBOW-An-open-trusted-and-secure-fog-computing-platform

### 4.3.5 NextCloud Repository

RAINBOW data exchange platform (Nextcloud) applies technological and organizational measures to secure processing of personal data against publishing to unauthorized persons, processing in violation of the law and change, loss, damage or destruction.

- **Information security**: SSL (Secure Socket Layer) certificates are applied. In order to ensure the appropriate level of security, the password for the account will exist on the platform only in a coded form.
- **Options for reading data**: the platform offers the possibility to make data available in a read-only or downloadable format, hindering the access to information by unauthorized users.
- **Back-up policy**: complete and redundant back-ups are done every week. Moreover, every time a modification is done an older version is saved.
- **Accidental deletion or modifications**: in case of a catastrophic event that implies the partial or complete deletion of the data sets, the data from the most recent back up will be automatically restored (back-up won't be older than 60 minutes). In case of accidental deletion or modification only the most recent document will be restored, so in case of accidental changes or deletion data can be easily recovered.
- **Deletion or modification of data by users**: only administrators have the rights to delete or modify the information included in the datasets.
- **Terms and conditions**: the Nextcloud platform have specific terms of use and conditions that have to be accepted by all users of the platform.

## 4.4 Access Data and Sharing

The accessing and sharing of data is firstly ruled by two documents : the non-disclosure agreement, which stipulates under which conditions transmitted information between the project partners is deemed confidential and must not be further disseminated; and the Description of Action (DoA) which stipulates the dissemination level of each deliverable. Moreover, the project consortium will comply with the FAIR (findable, accessible, interoperable and reusable) (European Commission, 2016) guidelines of the H2020 programme.

The data necessary to successfully complete the project Work Packages (WPs) will be shared without any restrictions amongst the WP partners either via internal repositories or direct communication. Public data will be made available at the project's website, the Royal Holloway data repository (Royal Holloway, n.d.-a) or other repositories, as appropriate. Users will be made aware of this data primarily through research publications, patent applications, dissemination activities, invited talks, social networks and the project website. Data will be made available to the project consortium as soon as it is available; to standardisation bodies when required; and to the public at the due date of the derivable, and, in case a research publication is based on that, as soon as the paper is submitted (if submission is anonymous, this will be postponed). If access to confidential data is necessary by the public, restrictive measures will be put in place.

# 5 DMP Template

The first draft of the RAINBOW DMP template is presented in Table 3-1.

The DMP template is available to the Project participants to handle the information related to the data management issues, such as:

- what types of data are generated/collected,
- what are the access/sharing restrictions of the data, if any,
- what format is used for storing the data, and
- what are the long-term storage and backup solutions and the suggested lifetime.

Additional information may be added in case that additions/modifications to the existing table fields will be required.

As FAIR methodology instructs, the goal is to make sure all relevant data in the project are collected in a structured and documented way, stored in a long-term accessible format, enriched with appropriate meta data allowing them to be found and make them usable for a longer period.

Regarding the nature of the possibly provided data, as already stated above, the RAINBOW Consortium does not foresee to produce huge amounts of data intended be exploited by other research activities, nor the need to collect re-usable raw data during the project life cycle.

The project activities most likely to create interesting data to be put at disposal of the scientific and research community can be easily identified with the planned RAINBOW demonstrators.

The project demos have been planned to be organized within the works of WP6, "RAINBOW Industrial Demonstrators and Performance Evaluation", and will described in the deliverables D6.7 and D6.13 "Validation Results, Performance Evaluation and

Adoption Guidelines", related to the first and second demonstration and evaluation phase, respectively. These two reports will document the evaluation results gathered from the execution of the RAINBOW demonstrators and the operation of the testbeds, which constitute the groups of deliverables D6.2-D6.6 (first phase) and D6.8-D6.12 (second phase).

All RAINBOW partners will contribute in defining and gathering the demo data set that will be available as needed.

Finally, the RAINBOW DMP will take into account the needs of protection for the relevant data, in case of any specific requirement.

As already stated, RAINBOW DMP approach will comply with the rules as suggested in the "Guidelines on FAIR Data Management in Horizon 2020", «the DMP is intended to be a *living document*», which is expected to mature during the project and to be finalized and delivered as a final version at the end of the project. Thus, this template must be intended as a first draft, which will undergo some changes throughout the project's lifetime to better fit the data and outcomes that could be generated during the RAINBOW research activities.

The template below presents four main fields, each one characterized by a variable number of different categories that is described in the second column.

*Table 5-1: RAINBOW DMP Template*

| **Data type definition** | |
|---|---|
| **Name** | Univocal identifier of the considered data |
| **Format** | Data format, measuring unit, typical order of magnitude |
| **Data Generation** | Description of the type of input used to generate the data |
| **Data Collection** | Description of the complete methodology and tools used for data collection |
| **Replicability tools definition** | |
| **Repository** | Description/location of the available data |
| **Emulation Tools** | Description/location of possible emulation tools useful for replicating the data |
| **Standard Software Interfaces** | List of the standards used to promote results replicability |
| **Extensions to Standard Interfaces** | Extensions to the above standards as developed during the project |
| **Data sharing** | |

| | |
|---|---|
| **Data Dissemination Strategies** | Promotion/information on the collected data characteristics and ways of exploitation |
| **Shareability Restrictions / Related Information** | Conditions and consequent actions in the case of need for protection of the relevant data |
| **Data maintenance** | |
| **Quality Consistency** | Constraints determining the quality/currency of the collected data |
| **Data Backup** | Consistent location of the data, including previous releases |

# 6 Ethics and Legal Compliance

The Rainbow consortium is aware of ethical, privacy, copyright and data protection issues that might be raised by the activities to be performed in the scope of the project. Subsequently, this section revolves around ethical and legal compliance issues as for instance the consent for data preservation and sharing, protection of the identity of individuals and how personal data will be handled to ensure it is stored and transferred securely as well as copyright and intellectual property rights (IPR) issues.

In light of the aforementioned the consortium deemed appropriate to peruse on the one hand the legal framework that governs the activities of the project i.e a) the European Union ePrivacy Directive 2002/58/EC, b) the Directive 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data thus repealing Council Framework Decision 2008/977/JHA and c) the (EU) Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data thus repealing Directive 95/46/EC and on the other the ethical guidelines for research projects in the EU under Horizon 2020 which must comply with ethical principles and relevant national and EU legislation i.e. a) Horizon 2020 Rules for Participation: Ethics Reviews (Article 14), b) Horizon 2020 - Regulation of Establishment: Ethical principles (Article 19), c) Model Grant Agreement: Ethics (Article 34) and d) the document "H2020 Programme Guidance How to complete your ethics self-assessment,

The Rainbow consortium will act in accordance with the above-mentioned legislation and Horizon's 2020 ethical principles as it pertains to any individual that might be involved in the project either as a participant or not.

## 6.1 Privacy Control and Informed Consent

The Consortium for the following activities : a) organization of workshops, b) the establishment of contact points, c) community building activities, d) creation of a network of potential users, e) use of questionnaires and f) personal interviews will provide to all individuals participating, information on the procedures undertaken and how their data are being handled so that every individual will be able to have access and control over its own data as well as to provide its consent. This is why an informed consent form (see **Error! Reference source not found.**) will be handed out to any individual participating in RAINBOW activities which may lead to the collection of data.

## 6.2 Confidentiality

Rainbow partners must retain any data, documents or other material as confidential during the implementation of the project. Further details on confidentiality can be found in Article 36 of the Grant Agreement along with the obligation to protect results in Article 27.

## 6.3 Personal Data

Personal data which will be collected within this project, will only be stored, analysed and used anonymously. The individuals will be informed about the use of the information collected by them and will have to agree to the data collection while providing their approval in the form of written consent. The identity of any individual interviewed or in any other way engaged in the project (e.g. by email, correspondence, newsletter) will be protected by this anonymization of the data.

## 6.4 Intellectual Property Rights (IPR)

Rainbow undertakes to ensure that data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and, in the EU, as well as with the H2020 rules. Raw data collected through the interviews from external to the consortium sources may be available to the whole consortium or specific partners upon authorization of the owners. This kind of data will not be available to the public. Concerning the results of the project, these will become publicly available based on the Access Rights as described in the Consortium Agreement.

# 7 Conclusions and Next Steps

Since this document is released in an initial stage of the project, it can only include an indicative description of the data, based on what is expected to be generated by each partner. Due to the dynamic nature of the project, the DMP is intended to be a living document that can be updated during the project lifetime, in order to be able to reflect important changes to the project. Moreover, D7.1 includes the core version of the DMP where all ethics requirements are addressed.

A complete and thorough description of the DMP will be provided at the end of the project (under one document), once all data has been collected, processed or generated and thus more detailed information can be provided. Then, the final version of the DMP will be released.