IMP-PCLOUD project

Data Management Plan

Version 1.0

Authors: Rafael J. Segura-Sánchez (rsegura@ujaen.es), Carlos J. Ogayar-Anguita (cogayar@ujaen.es)

Grant No.: <PENDIENTE>

Project Start Date: <PENDIENTE>

Project Duration: <PENDIENTE>

Change control

Version	Date	Author	Organization	Description/Comments
1.0	2024/01/17	Rafael J. Segura-Sánchez	University of Jaén	First draft

Contents

Summary

This is the preliminary version of the Data Management Plan (DMP) for the project IMP-PCLOUD. It contains a description of the data that will be produced during the execution of the project, as well as how this data will be used, stored and published.

Since this project is publicly funded, and according to the Open Science principles[1], all the data and scientific publications that are not protected by copyright laws will be publicly accessible, applying the guidelines described in this document.

This document is not final, as it might be modified during the research lifecycle.

---

[1] https://www.unesco.org/en/open-science/about

1. Data description

During the execution of the IMP-PCLOUD project, two different types of data will be produced, namely point cloud datasets and research papers.

1.1. Point cloud datasets

As this project is aimed at dealing with massive point clouds, one of the tasks will be related to the generation of these datasets, either by collecting existing data publicly available or using datasets already created by this research team in the past. In addition, new point clouds will be obtained as part of this project, using 3D scanning systems available at our institution.

These point clouds will be handled in digital form, using standard file formats like LAS and LAZ[2], and additional metadata describing them will be also produced. The dataset files are expected to be of very large size, given the amount of points contained in them. Therefore, the storage requirements for this data have to be carefully considered. In addition to this, data will also be generated regarding the spatial organization of the data, version control data, as well as other data that needs to be created as a result of the development of the different algorithms. This data will be stored both in databases and file systems.

As this project is focused on point cloud datasets from natural and human-created locations, no ethical issues related to data privacy and personal intimacy are expected to appear. For the case of new datasets, the locations to be scanned will be carefully selected in order to prevent these conflicts from appearing.

1.2. Research papers

As a result of our developments and experiments, we expect to be able to submit and publish a series of scientific papers in specialized journals. These papers will contain descriptions of the completed work, the algorithms used and/or developed and the results obtained, as well as comparisons with other algorithms currently published and in use by the industry and the scientific community. No matter how these papers are published, it will be necessary to provide public access to a version of these documents.

2. Data lifecycle

This section explains how the data described above will be dealt with, not only during the project execution, but also once it has finished.

---

[2] https://github.com/ASPRSorg/LAS

In application of the FAIR principles[3], a key element for the management of the data will be the infrastructure provided by the University of Jaén; to be precise, the RUJA[4] repository provided by this institution to publish research results with Open Access (OA). All the information published in this resource, unless stated otherwise, is available under a Creative Commons (CC by-nc-nd) license[5].

2.1. During the project execution

As the project execution develops, different point cloud datasets will be obtained. For this purpose, two strategies will be implemented:

- The first one will consist of reusing existing datasets to create new ones; this will include publicly available datasets, as well as datasets created by ourselves in the past. These datasets will be edited if necessary, in order to suit the needs of this project in terms of file format, size and metadata.
- In order to produce totally new datasets, 3D scanning systems will be used on different locations, both natural and human-created, so as to include as many different features as possible.

All these datasets will be stored using standard file formats like LAS and LAZ, plus additional files related to spatial organization, version control data and others. Proper metadata will also be added, in application of the FAIR principles.

These datasets will be kept private during the development of our experiments, but once they are finished and the research results are submitted for publication, they will be made available through RUJA, so that reviewers can verify our achievements.

Regarding research papers, once they have been accepted for publication, they will be made publicly available, either through OA journals or the institutional repository RUJA. In any case, copyright infringements will be avoided by checking the OA policies of the journals using resources like Sherpa Romeo[6]. Together with the paper, corresponding metadata will be added to the corresponding repository item, in order to make it easily findable and accessible, in application of the FAIR principles.

2.2. After the end of the project

Once the project execution has reached its deadline, the resources uploaded to RUJA will remain publicly available indefinitely, as this repository is intended to be permanent in time.

---

[3] https://www.go-fair.org/fair-principles
[4] https://ruja.ujaen.es
[5] https://creativecommons.org/licenses/by-nc-nd/4.0
[6] https://v2.sherpa.ac.uk/romeo

Those datasets and papers that at the end of the project have not yet been published will be treated in the same way as described above, i.e. the efforts for publication will continue until success, and once they have been published, both the datasets and the published papers will be uploaded to RUJA and made publicly available.

3. Data responsibles

The data collected and/or created during the execution of the project will be under custody of the research team while they are used for experimental purposes. Every person in the team could be assigned the custody of some data, as long as it is related to the project-related tasks they have been assigned.

The University of Jaén is responsible for RUJA, in terms of keeping the site working and preserving the uploaded data. This repository assigns a permanent URI to each uploaded work and complies with the Open Archives Initiative (OAI) standards[7], in order to make the data comply with the FAIR principles.

---

[7] https://www.openarchives.org