AIPlan4EU

Bringing AI Planning to the European AI On-Demand Platform

D1.3. Open Research Data Management Plan

June 30th 2021

Project funded by the European Commission within the Horizon 2020 Programme		
Dissemination Level		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	□
CL	Classified, as referred to in Commission decision 2001/844/EC	□

Deliverable number:	D1.3.
Deliverable name:	Open Research Data Management Plan
Work package:	WP1. Management
Lead WP:	FBK
Lead Task:	FBK

Contents

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main change
06/05/2021	V1	Paola Baruchelli (FBK)	Initial draft
07/06/2021	V2	Andrea Micheli (FBK)	Completed draft
11/06/2021	V3	All Partners	Shared with all the partners
30/06/2021	V4	Andrea Micheli (FBK)	Final revision

Abstract

Automated Planning and Scheduling is a central research area in AI that has been studied since the inception of the field and where European research has been making strong contributions over decades. Planning is a decision-making technology that consists in reasoning on a predictive model of a system being controlled and deciding how and when to act in order to achieve a desired objective. It is a relevant technology for many application areas that need quick, automated and optimal decisions, like agile manufacturing, agrifood or logistics. Although there is a wealth of techniques that are mature in terms of science and systems, several obstacles hinder their adoption, thus preventing them from making the footprint on European industry that they should make. For example, it is hard for practitioners to find the right techniques for a given planning problem, there are no shared standards to use them, and there is no easy access to expertise on how to encode domain knowledge into a planner.

The AIPlan4EU project will bring AI planning as a first-class citizen in the European AI On-Demand (AI4EU) Platform by developing a uniform, user-centered framework to access the existing planning technology and by devising concrete guidelines for innovators and practitioners on how to use this technology. To do so, we will consider use-cases from diverse application areas that will drive the design and the development of the framework and include several available planning systems as engines that can be selected to solve practical problems. We will develop a general and planner-agnostic API that will both be served by the AI4EU platform and be available as a resource to be integrated into the users' systems. The framework will be validated on use-cases both from within the consortium and recruited by means of cascade funding; moreover, standard interfaces between the framework and common industrial technologies will be developed and made available.

Keywords definitions

The definition of reference are the ones reported by the H2020 online manual[1].

Key word	Definition
Data Management Plan (DMP)	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).
Digital Object Identifier (DOI)	Permanent identifier which should be a persistent link to the published version full text or abstract (if article is pay per view) or to the final manuscript accepted for publication (link to article in repository).
License	Legal conditions under which an item or piece of knowledge being transferred is provided
Metadata	Information about the research data, structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection.
Open Access (OA)	Open access can be defined as the practice of providing on-line access to scientific information that is free of charge to the reader.
Open Access to Research Data	The right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement.
Personal Data	'Personal data' are defined extremely broadly and include 'any information relating to an identified or identifiable natural person'. An 'identifiable natural person', or 'data subject', is 'one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' (Article 4(1) GDPR).
Repository	Digital archives collecting, preserving and displaying datasets, related documentation and metadata.
Research Data	Refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form. Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge.

Table 1. Key words definitions

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/index_en.htm

Executive summary

This deliverable is the first version of the Data Management Plan (DMP) for the AIPlan4EU project submitted at M06 (June 2021 – D1.3). The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

The present document provides an analysis of the data management policy and strategy agreed by the partners to the dataset generated within the AIPlan4EU project, identifying the main data to be generated and collected by the project, outlying the handling of research data during the project and how part of them will be openly shared.

AIPlan4EU partners will internally use this document as a guide on data management collecting common practices and policies.

1 References

The basis for the construction of this document is both the official AIPlan4EU documentation as well as the EC references listed below:

- AIPlan4EU Grant Agreement and its Annexes;
- H2020 Annotated Model Grant Agreement - Open access to research data[2];
- Guidelines to rules on Open Access to Scientific Publications & Open Access to Research Data in Horizon 2020[3];
- Guidelines on FAIR Data Management in Horizon 2020[4];
- Template for the Data Management Plan[5];
- OpenAIRE Research Data Management Briefing Paper[6].

The following table reports the main reference for the DMP extracted by the indicated documents.

Document	Access[7]	Availability	Relationship with AIPlan4EU DMP
AIPlan4EU Grant Agreement	Confidential	Participant Portal	• Article 27 details the obligation to protect results (27.1) and of providing information on EU funding (27.3) • Article 29 details the obligation to disseminate results, defines open access to research data (29.3) as well as the obligation to provide information on EU funding (29.4) and to exclude Commission responsibility via a disclaimer (29.5) • Article 36 details confidentiality obligations • Article 37 details security-related obligations • Article 39 details obligations to protect personal data
AIPlan4EU Grant Agreement – Annex 1 Part A	Confidential	Participant Portal	WP1 - T1.4 Ethics, IPRs and Data Management

[2] Version 5.0, October 2017 - https://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf

[3] Online version: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

[4] Version 3.0, April 2016 - https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[5] Version 1.0, October 2016: http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tploa-data-mgt-plan_en.docx

[6] 5 April 2017, https://www.openaire.eu/briefpaper-rdm-infonoads

[7] *Confidential*: limited to Consortium, European Commission, appointed external evaluators and other EU bodies; *Consortium*: originally conceived as consortium but can be made available to European Commission, appointed external evaluators and other EU bodies if necessary; Public: public and fully open availability.

D.1.1. Project Communication Platform and Archive	Public	AIPlan4EU Website	The deliverable further defines the project target stakeholders as well as IPR plan and policies of the project. The deliverable will detail the communication materials including visual identity and basic project rules for communication tailored to different target audiences (project logo, PowerPoint presentations, social network posts)

Table 2. Relation to the project key documents and deliverables

2 Definitions

2.1 Privacy

All partners are committed to comply with applicable rules and regulations, in particular the compliance with the GDPR[8] and with the FAIR principles.

The GDPR prescribes that management of personal data is the responsibility of each legal actor in possession of such, and that each such actor is required to have a data protection officer (DPO).

Each partner organization is thus obliged to have a data protection/data management policy for management of personal data (Local DMP), including routines for reviews and updates. Each partner will, when required according to the GDPR, do DPIAs (Data Protection Impact Assessments), under due supervision of the organization DPO. Ethical aspects of management of personal and personal sensitive data

Should absence of required local policy documents be identified, the partner will be encouraged to rectify. Should a conflict result, the consortium conflict resolution routine as described in the agreements will be used to solve it.

2.2 FAIRness

A part of the Open Research Data (ORD) pilot is concerned with making data management FAIR, that is: datasets are findable, accessible, interoperable and re-usable. The principles imply no restrictions about specific technology, standard, or implementation-solution. AIPlan4EU is a network with 16 partners including research centers, universities, SMEs and companies, and the technologies, standards, and solutions for assuring compliance with the FAIR principles will span a very broad range.

All AIPlan4EU partners that are legal owners of data must assure that local data management plans or policies (Local DMP) that cover the FAIR principles are in place.

3 DMP management and update

The project's first development and future updates mainly rely on the collection of Data Management forms filled out for each project datasets by project partners responsible for producing data (see Annex 1). Two formal revisions to be

[8] Regulation (EU) 2016/679 (General Data Protection Regulation) in the current version of the OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018

submitted to the European Commission are foreseen in M18 and M48. Nevertheless, the document will be open for constant update and further development all along the project duration.

Official versions will be stored on the project online repository as PDF files. An editable word copy of the latest version will also be stored to facilitate revision and update of the already identified datasets and policies. Should a new dataset be identified along the project implementation, Partners can submit a new form to the coordinator. FBK will then oversee updating the document and its annexes, uploading them to the repository and notifying the consortium through the project mailing list system.

4 Data Summary

In this section, we present the data the project will generate and use and indicate the purposes and expected usages of such data.

4.1 AIPlan4EU datasets

The AIPlan4EU consortium encompasses 7 partners that are in charge of providing use-cases in different application sectors; in addition, part of the cascade funding will be used to recruit additional use-cases for the planning technology that the project will develop. Each use-case is constituted by a textual description as well as additional data that will be needed for the experimentation and evaluation of each use-case. In this respect, AIPlan4EU will produce the following datasets.

- Agriculture use-case:
  - Historical data on agricultural practices and measures of soil compaction and resource consumption
  - Models of soil compaction and resource consumption
  - Machine availability and requirements
- Flexible manufacturing use-case:
  - Historical data of operations for different kinds of production
  - Models of factories and their capabilities
  - Models of resource consumption
- Underwater robotics use-case:
  - Historical telemetries
  - Models of resource consumption
  - Models of underwater structures
  - Models of robotic capabilities and timings
- Space use-case data:
  - Historical telemetries
  - Models of space-exploration robots
  - Models of resource consumptions
- Logistics use-case data:
  - Historical telemetries
  - Models of space-exploration robots
  - Models of resource consumptions
- Fleet-management use-case data:
  - Policies and rules for the vehicles
  - Models of battery consumption
  - Historical data and performance metrics
- Consumer product experimentation use case data:
  - Composition of example products

- Historical data and performance metrics
- Historical telemetries

- Open-call use-cases will provide additional data that needs to be managed within the project. At this stage we cannot anticipate the kind of data that will be provided, but we will update this document accordingly as soon as the additional use-case will be selected and engaged.

In addition to use-case data, the project will also produce software artifacts, reports and papers. Software artifacts are of two kinds: the "Unified Planning Framework" (UPF) and the various "Technology-Specific Bridges". Concretely, the UPF is a software library that is released as Open Source under the terms of the Apache 2.0 License and is freely available at https://github.com/aiplan4eu/upf. The TSBs will be developed in WP5 and will serve the technologies that are peculiar to each use-case. The TSBs licenses will be "as open as possible" considering IPR issues and the exploitation plan of the project partners. Both the UPF and the TSBs will also be developed and archived on GitHub[9] and will be delivered as Public and Confidential project deliverables respectively (D3.4 and D5.4). Academic papers will be published in reputable journals and conference proceedings in open-access form, and we will make sure that a suitable Digital Object Identifiers (DOIs) is available for each produced paper. Papers might use anonymized and/or reduced dataset collected within the project: in these cases, we will use the Zenodo[10] platform to safely archive these datasets for preservation, for the reproducibility of the experimentation and so that the data is findable and persistently citable.

4.2 AIPlan4EU data purpose and utility

The purpose for the data generation and collection within the AIPlan4EU activities is threefold.

1. *Characterization and reproduction of the planning problem identified for each use case.* Each use-case involved in the project will bring the data needed to encode a planning problem. This data is almost always technical and non-personal but might be proprietary of the companies owning the use-cases.
2. *Experimentation and validation of each use case.* Historical data and performance metrics are needed to perform the evaluation of the developed solutions and technologies that will be developed within the AIPlan4EU project.
3. *Categorization of use-cases according to the kind of planning problem they entail and association with the most appropriate solution*. Problem and use-case characteristics will be analyzed (manually and automatically) to recommend the best planning technique given a use-case and to develop portfolio-based solutions.

Detailed objectives for each dataset to be generated will be identified by the partner submitting the dataset identified in Annex 1 table. Each partner will submit the form to the coordinator when a dataset is ready to be published on the repository. Aside from the utility for the AIPlan4EU project, the data to be developed may be useful to several external entities and stakeholders who have been already identified in both the Description of Action and the Dissemination Plan. These mainly include Scientific community, Education community and companies.

In order to maximize the dataset interoperability, management and re-use, AIPlan4EU consortium agreed to use, when possible and compatibly with the industrial requirements and data availability, formats that are non-proprietary, unencrypted, uncompressed and commonly used by the research community.

4.2.1 Data Re-use

For the development and experimentation within the project, we will use existing planning problems datasets that are commonly available in the literature. Examples include the International Planning Competition collection of

[9] https://github.com/aiplan4eu
[10] https://zenodo.org/

benchmarks[11] and existing industrial problems such as HSP[12]. These benchmark and example problems will be used to evaluate the development progress on WP3 and will serve as prototypical examples to guide and inform the development of Technology-Specific Bridges in WP5.

4.3 FAIR data

4.3.1 Making data findable, including provisions for metadata

As mentioned, we will use GitHub and Zenodo for archive and make findable the software artifacts and the datasets used in academic experimentation; each academic dataset will have a Digital Object Identifier (DOI) so that it is findable and persistently citable. Zenodo, in fact, assigns a versioned DOI and gives the possibility to update the shared material. AIPlan4EU will therefore use the linear versioning supported by Zenodo[13].

Each project record will be annotated with metadata that includes the identifier of the data it describes and that are indexed in a searchable resource to increase data findability.

In addition to this, complete project documentation will accompany the data to allow correct data interpretation and eventual experiment reproduction. This will include:

- Dataset overview – number of sub-datasets, if any; status of documented data (complete or in progress); eventual plan of future update
- Methodological information – methods used for experimental design, data collection and data processing; instruments and software used; experimental conditions; quality assurance procedures performed on data.

4.3.2 Making data openly accessible

AIPlan4EU will follow the "as open as possible" principle for its data, however some datasets arising from the use-case might be sensitive and/or covered by IPR. Hence, three different levels of confidentiality are considered:

1. *Confidential to a partner*. Dataset that contains data that cannot be shared by the owner partner (for instance data protected by partner's internal regulations) or be shared only between a limited number of partners in the consortium (for instance non complete datasets or datasets that must be generated in a collaborative experiment). Data expected to be included in patent applications, if any, will be shared between partners at this level of confidentiality.
2. *Confidential to the consortium*. This option is applied for data not yet published in peer reviewed scientific papers or that are intended to be published before they are made publicly available.
3. *Public*. This option is applied to every other dataset.

Datasets confidential to partner or to the consortium will be safely stored by the developing partner and in the project shared repository. Public data will be shared via one or more of the following options:

- the AIPlan4EU website;
- the GitHub and Zenodo services as described above;
- the AI On-Demand Platform.

4.3.3 Making data interoperable

Most of the data produced and used within the project consists of planning models encompassing the use-case knowledge needed to encode and ultimately solve the planning problems. The AIPlan4EU project will develop a new

[11] https://ipc2018.bitbucket.io/

[12] https://doi.org/10.1609/aaai.v33i01.33017675

[13] https://help.zenodo.org/#versioning

and innovative way of encoding this knowledge using a convenient software library (i.e. the UPF) and most of the models used within the project will be encoded using the UPF library. However, to facilitate the interoperability and to leverage existing planning models, the UPF library (developed within WP3) will offer conversions to and from popular planning languages used in the literature such as PDDL and ANML.

Moreover, the project will develop an ontology of planning problems, techniques and methods that will be aligned and integrated with the one developed by the AI On-Demand Platform and synchronized with the partners of the StairwAI ICT49 project.

4.3.4 Data re-use and licensing

Public data will be made available for re-use. To avoid any potential doubt, the consortium will attach specific licenses to the deposited data to define all conditions under which the work is provided under either open or restricted access. We will consider and favorably prefer the Apache 2.0 license for software artifacts and the Creative Commons licenses for documents and models. At the time of writing, we are still in the process of understanding which of the data generated within the use-cases will be published and which will be kept confidential. Successive versions of this document will detail these aspects.

5 Allocation of resources

The AIPlan4EU does not allocate a specific budget for these actions. The only costs foreseen for data management are personnel costs to work to establish and implement the data collection and analysis activities. These activities are part of WP2 – Requirements, focused on the collection and analysis of use case requirements and related data, and WP6 – Evaluation, focused on the experimentation and validation of the use cases. Moreover, academic papers stemming from the project work will be published as open-access and will be funded by the project grant.

FBK as project coordinator is in charge of the DMP from both the scientific and technical perspective, this includes the first version release as well as the regular update. Validation and registration of datasets and metadata, as well as backing up data for sharing through open access repositories is the responsibility of the partner that generates the data in the WP. Each partner will identify a specific responsible person for each dataset. Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator. Each partner should respect the policies set out in this DMP.

In line with Grant Agreement (art 29.1), each partner should give at least 45 days prior notice to the other partners before disseminating/publishing data.

6 Data security

AIPlan4EU is in charge of backing up data that will be openly shared through the project website and to maintain the project shared folder. The latter is hosted by the Google cloud services and is regularly and automatically backed-up. Moreover, the data is kept secure by the same cloud service.

7 Ethics aspects

The AIPlan4EU project will not collect, elaborate and/or store personal and human data. As described above the datasets involved in the experimentation and validation of the use cases will be not related to humans.

Ethical concerns of the AIPlan4EU project are related to AI and robotics solutions that should have[14]:

- *Social impacts* on labor market and economy, inequality, privacy, human rights and dignity. To reduce this risk the reference will be:
  - the OECD[15] first principle of inclusive growth, sustainable development and well-being states that AI should be developed in a way that reduces 'economic, social, gender and other inequalities';
  - the second OECD principle, which states that AI systems should respect diversity and include safeguards to ensure a fair society, however detail on how this can be achieved is lacking.
  - the EU ethics guidelines[16] that elaborate that equality is a fundamental basis for trustworthy AI and state that AI should be trained on data which is representative of different groups in order to prevent biased outputs.
- *Psychological impacts* related to the human-robot relationships that must – according to EC guidelines – be carefully monitored and considered.
- *Financial system impacts* in terms of manipulation and collusion and the need to build in accountability as a key principle and states that "organizations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning" (OECD). It is likewise a core principle of the EU ethics guidelines.
- *Legal system impacts* related to the need of new laws to regulate the use of AI and robotics and possible injuries caused. The issues of liability, safety and security are explicitly addressed by the EU in both its Communication[17] and ethics guidelines.
- *Environmental impacts* in terms of increasing energy demand, natural resource use, waste production, but always considering also the positive outcomes for the planet.
- *Impacts in trusts* including aspects such as fairness, transparency, explainability, accountability, and control where human-in-the-loop approach is proposed.

In line with these principles, the AIPlan4EU partners will evaluate case by case and plan contingency actions[18] to reduce negative impacts and risks, updating this document. In particular, we will constantly supervise the developments of the open-calls to update the ethical considerations if additional use-cases will pose issues in these respects.

[14] European Parliament (2020), *The ethics of artificial intelligence: Issues and initiatives – Panel for the Future of Science and Technology*, EPRS European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020

[15] OECD (n.d.) OECD initiatives on AI [online] Available at: http://www.oecd.org/going-digital/ai/ [Accessed 12 May 2020]

[16] European Commission (2019), *High-level expert group on artificial intelligence: Ethics guidelines for trustworthy AI*. Brussels. - https://ec.europa.eu/futurium/en/ai-alliance-consultation

[17] European Commission (2018). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe*. https://ec.europa.eu/digital-singlemarket/en/news/communication-artificial-intelligence-europe

[18] Reference will be here IEEE recommendations about new governance frameworks, standards, and regulatory bodies (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems - https://standards.ieee.org/industry-connections/ec/autonomous-systems.html [Accessed 13 May 2021])

Annex1: AIPlan4EU dataset form

Responsible partner	
Other partners involved	
Goal and purpose	
Relation to AIPlan4EU project objectives	
Related to AIPlan4EU task(s)	
Dataset type	
Data Origin	
File format(s)	
Expected volume data	
Expected time of delivery	
How data will be acquired and processed	
Data Repository	
DOI	
Permanent link	
Metadata description	
Restriction on sharing	
Copyright & IP issues management	
Duration	
Licensing	
Resourcing	
Long term value	
Backup procedures	
Ethical issues management	