H2020 Grant Agreement No. 101004719 (ORP)

PROJECT TITLE:	OPTICON RadioNet Pilot - ORP
STARTING DATE	01/03/2021
DURATION:	48 months
CALL IDENTIFIER:	H2020-INFRAIA-2018-2020 submitted for H2020-INFRAIA-2020-1
TOPIC:	INFRAIA-03-2020
	Pilot for a new model of Integrating Activities

OPTICON RadioNet Pilot

Deliverable 2
Data Management Plan

Due date of deliverable: 2021-08-31

Actual submission date: 2021-09-15

Leading Partner: CNRS

Document information

Document name:	**Data Management Plan**
Type	**Open Research Data Pilot**
WP	**WP1 – Management**
Version date:	**30/07/2021**
Authors (Institutes)	**Hanno Holties (ASTRON)**
	Hélène Dworak (CNRS),
	Jean-Gabriel Cuby (CNRS),
	John Davies (UKATC/STFC),
	Lukasz Wyrzykowski (UNIWARSAW),
	Robert Beswick (UNIMAN),
	Tim Morris (UDUR)
	With the participation of other WPL and EMT members

Dissemination Level

	Dissemination Level	
PU	**Public**	**X**
PP	**Restricted to other programme participants (including the Commission Services)**	
RE	**Restricted to a group specified by the consortium (including the Commission Services)**	
CO	**Confidential, only for members of the consortium (including the Commission Services)**	

History of Change

Version	Date	Change	Page
1	31.07.2021	First draft	All
2	31.08.2021	Initial DMP	All
3		Mid-term DMP	
4		Final DMP	

Index

I.   Introduction

This document (deliverable D2) describes the Data Management Plan (DMP) for the OPTICON RadioNet Pilot (ORP) project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement number 101004719. The purpose of the DMP is to set out the main elements of ORP data management policy for the data-sets generated by the project. It draws on the Horizon 2020 FAIR Data Management Plan (DMP) template (Version 3.0, 26 July 2016) and belongs to the Open Research Data Pilot.

The DMP provides a summary of the expected type and format of data that the project will generate (section III), and outline data management in practice as well as the policy in relation to data resources (section IV), security, ethics and confidentiality (section V to VIII). This DMP is a living document and will be updated at regular intervals during the project.

II.   Project summary

The OPTICON-RadioNet Pilot (ORP) brings together the well-established ground-based astronomy community, in an effort to support and develop seamless access to radio and optical facilities in an efficient, coordinated and future-looking programme. It offers access to an unrivaled set of major and specialised observatories across Europe (and around the world) covering the optical, infra-red, sub-mm and radio wavebands to open the way to new discoveries.

The ORP mission is to facilitate astrophysical discoveries with a comprehensive set of research infrastructures in the related domains of optical and radio astronomy with a harmonised access procedure, unified data interfaces, and improved services toward scientific excellence.

III.   Data Summary

The main purpose of the Data Management Plan (DMP) is to describe the data management life cycle for the data to be collected, processed and/or generated by the ORP project. It also aims to provide a framework to support the European Commission's goals for Open Access regarding publications, scientific and technical results and raw data resulting from activity supported by the EC Grant Agreement.

It is a requirement of Horizon 2020 grants that publications resulting from the grant should be made in an Open Access journal unless there are compelling reasons why this should not be done. Outputs from the ORP activities may be grouped into a number of different types:

> 1/ Scientific data resulting from the Transnational Access programme
> 2/ Scientific publications resulting from the Transnational Access programme
> 3/ Technological or software research and development.
> 4/ Technical data and publications resulting from Management and JA 1-2-3-4
> 5/ Technical and personal data resulting from the CTAC

As part of making research data findable, accessible, interoperable and re-usable (FAIR), the DMP will identify for each category:

- Data set reference and name: identification of what data will be collected, processed and/or generated
- Data set description: description of the data set
- Standards and metadata: explanation of the methodology and standards that will be applied
- Data sharing: specify whether data will be shared/made open access or not. How will data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- Archiving and preservation (including storage and backup): explanation of how data will be curated and preserved (including after the end of the project)

The data collected and generated by the ORP project will be useful for astronomers, universities, science students, etc. By sharing the data, the project will contribute to additional scientific discoveries by re-use of data taken for other purposes.

IV.   FAIR data

1.   Scientific data

a.   Data set reference and name

The Scientific data set includes all scientific data resulting from the Transnational Access (TA) and offered under the Virtual Access (VA) programs.

b.   Data set description

The data sets collected from the optical and radio infrastructure include digital research data (i.e. raw data and data calibrated via pipelines), and metadata associated with observations and/or data processing procedures. These data and metadata will be used for scientific research and publications, as well as enabling peer-review on scientific publications related to the program results.

c.   Standards and metadata

In most cases, the data are stored in common data formats such as FITS or MeasurementSets, which are data formats that include rich metadata describing the content and can be easily read by most astronomical data reduction packages. New standards for adaptive optics data are being defined in the Pilot, to complement FITS.

d.   Data sharing

Raw data and data calibrated via pipelines, as well as metadata from observations, are kept in searchable archives which are maintained by the participating observatories. Access to the data is available upon request, in line with the open access provisions for scientific publications. Digital research data and appropriate metadata are stored, managed and made available in a long term and sustainable way that does not depend on project funding.
In almost all cases, the observed data become publicly available after a proprietary period. This is typically one year, with the clock starting at the time of observation or the time of the distribution of post-correlated/pre-processed FITS files to the PI, depending on the specific ORP TA facility.

Due to the extremely high data rates required for radio-interferometric observations only pre-processed data, resulting in reduced data volumes, such as correlated, beam-formed and/or compressed data can be stored. These correlated data are made available through facility archives.

In most cases, data require additional reduction steps performed by the PIs of the observing projects, often with support from Pilot TA centers. For data resulting from radio facilities and where applicable, ORP will encourage the facilities to make these data workflows public via recognized workflow/science data repositories such as Zenodo. The resulting figures and tables may be made publicly available by the PIs via the Vizier Data catalogue maintained at the Centre de Données Astronomiques, CDS, (Strasbourg, France), in addition to the material published in printed and electronic versions of international research publications.

Regarding the Canary AO testbed, Data will be shared through the AO virtual access portal to be developed by JA3.3.2.

Regarding TDA-CCS, the Time-Domain photometric images collected via or uploaded and processed via TDA-CCS are stored in the calibrated form, but the raw data will be deleted after 1 month. The time-series data which are results from the pipeline will be archived and made available throughout and beyond the Project.

e.   Archiving and preservation (including storage and backup)

Data generated for the ORP project is archived by the facilities that collaborate in the project. The archives are operated by facilities with a long-term perspective, all data will be available during and after the project´s lifetime.

These archives are accessible via web interfaces, most of which complying with the Virtual Observatory (VO) standards. These data and associated metadata are either directly downloadable or are accessible by interaction with the observatory staff or provided services.

User experimental data will be archived at user institutes and is the responsibility of the ORP facilities.
Data archives of the ORP infrastructures 4.1:

TA Name	Archive Address	Access conditions
EVN	http://www.jive.eu/select-experiment	Free after 1 year Exception on raw FITS files and pipeline output[1]
e-MERLIN	http://www.e-merlin.ac.uk/archive/	Free after 1 year
Effelsberg	http://www.mpifr-bonn.mpg.de/en/effelsberg	Free, upon request
LOFAR	https://lta.lofar.eu/	Free after 1 year
IRAM	https://www.iram-institute.org/EN/content-page-386-7-386-0-0-0.html	Free after 3 year For large programs this period is reduced to 1.5 years.
APEX	http://archive.eso.org/wdb/wdb/eso/apex/form	Free after 1 year
ALMA	http://almascience.eso.org/alma-data/archive	Free after 1 year
SRT	under development	N/A

TA Name	Archive Address	Access conditions
AAT	http://site.aao.gov.au/AATdatabase/aat/aat.html	Free after 18 months
CAHA	http://caha.sdc.cab.inta-csic.es/calto/ http://caha.sdc.cab.inta-csic.es/calto/jsp/searchform.jsp.	Free after 1 year
CFHT	http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/cfht/	Free after 13 months
ING	http://casu.ast.cam.ac.uk/casuadc/ingarch/query	Free after 1 year
LT	http://telescope.livjm.ac.uk/DataProd/	Free after 1 year
OHP	http://www.obs-hp.fr/archives.shtml	Elodie: Fully available but OHP property Sophie: Free after 1 year Mistral: Free after 1 year
TBL	https://tbl.omp.eu/instruments/archive-tblnarval/	Free after 1 year Exception on Raw data available upon request
Aristarchos	http://helmos.astro.noa.gr/opticon.html	Free, upon request
LCO	https://archive.lco.global/ .	Free after 1 year
NOT	http://www.not.iac.es/archive/	Free after 1 year
REM	http://ross.iasfbo.inaf.it/REMDBdev/public.php	Free after 1 year
TCS	http://research.iac.es/OOCC/iac-managed-telescopes/telescopio-carlos-sanchez/	Free, upon request
TNG	https://ia2.inaf.it/index.php	Free after 1 year

2. Scientific publications

a. Data set reference and name

Scientific publications resulting from facility access

b. Data set description

---

[1] For more information see: http://old.evlbi.org/user_guide/archive_policy.html

Scientific publications result from observing runs at astronomical observatories, which have been supported under the ORP Transnational Access programme. They consist of peer-reviewed publication(s), containing a complete description of the algorithms used and measurements of system performance, presentations at conferences, and press releases.

c. Standards and metadata

Scientific publications are in general stored in PDF format.

d. Data sharing

When applicable, consortium members must ensure Open Access to peer-reviewed scientific publications resulting from ORP TA. The authors must retain sufficient intellectual property rights to comply with the EC requirements of open-access. Publication fees are only reimbursable for full open-access publications.

For publications resulting from facility access, partners are encouraged to adopt the 'gold' model by publishing in dedicated open access journals or by selecting applicable open access options. When possible, we expect the Pilot results to be published as articles in astronomical and multidisciplinary journals, and in the form of conference proceedings. Each observatory maintains a list of publications generated with its infrastructure. This list is linked to the NASA ADS (http://adsabs.harvard.edu/ads_abstracts.html) database of astronomical publications.

Regarding the Canary testbed, where publications are not open-access, pre-prints will be posted on ArXiv (Green Open Access). Any data used in the preparation of the paper will be made available as part of the CANARY data repository.

e. Archiving and preservation (including storage and backup)

Publications will be archived by the publishing journal, as well as in existing institutional repositories where appropriate such as arXiv (https:arXiv.org) or France HAL (https://hal.archives-ouvertes.fr/).

3. Technological or software research and development

a. Data set reference and name

1/ Software research and development

2/ Auto-tuned AO control and AO Virtual observatory

b. Data set description

1/ Development of a new proposal submission tool, the framework for common data access and processing, and that for multi-wavelength/facility and time domain access. The software will be useful for all the radio and optical infrastructures wishing to harmonise their processes.

 2/ The implementation of the software routines and GUIs that are required to implement the auto-tuned AO control algorithms on the CANARY testbed. It also includes Data exploration and exploitation tools for telemetry, control and point-spread-function derivation.

c. Standards and metadata

1/The standard code is not defined yet, as there is still some evolution on which languages maybe used. However, ORP WPLs commit to use a code formatted to improve readability and consistency among the project.

2/ Python software routines will follow the PEP-8 python coding standard.

d. Data sharing

Every software will utilise open access principles and recognised open-software and workflow repositories for curation, such as Zenodo and Github, accessed with an appropriate DOI. Produced software and its documentation will be available under open license, for example, via a GNU General Public License or Apache 2.0, and made readily available to any interested party.

e. Archiving and preservation (including storage and backup)

1/ The software products will be made accessible via the repository, such as Github/Gitlab. Direct download from the repositories will be available and do not require additional tooling.

2/ Regarding the Auto-tuned AO control and the AO Virtual observatory, source code will be archived as part of the CANARY software/data repository, with copies held in Durham and Observatoire de Paris.

4. Technical data and publications

f. Data set reference and name

Technical data and publications from Management and JA1-2-3-4.

g. Data set description

It includes technical publications published as the result of actions undertaken within ORP, such as training manuals, reports, booklets, press releases, statements, slides, course notes and/or training videos that are delivered as part of workshops, schools, trainings and events organized in the frame of the project.

The technical data can be defined by all the infrastructures data gathered for analytical and mapping purposes.

h. Standards and metadata

All material delivered as part of the programme will be stored in standard file formats – e.g. Word, PDF, power point for notes/slides, and MP4 (TBD) for training videos.

i. Data sharing

Any technical publications arising from the joint activity work packages will be accessible on the ORP website (https://www.orp-h2020.eu/), under the section *'Resources'* and made freely available to all.

Technical data will be shared within the Consortium through the ORP Redmine, unless otherwise specified, data will be confidential and not publicly shared.

j. Archiving and preservation (including storage and backup)

Technical publications will be archived on the ORP website, hosted by JIV-ERIC (Netherlands), and also on the ORP Redmine, hosted by the Observatoire des Sciences de l'Univers Institut Pythéas (France). Both institutes have sufficient resources to ensure the accessibility of the data during the lifetime of the project and 5 years after.

Technical data will be stored on the ORP Redmine or within the facilities archives.

5. Technical and personal data resulting from the CTAC.

a. Data set reference and name

Technical and personal data resulting from the Common Time Allocation Committee (CTAC), a panel of 7 external scientific experts who provide a peer review process of all the proposals submitted to the TNA process for optical telescopes.

b. Data set description

All the documents resulting from the CTAC process: Call documents, Proposal forms with personal data and scientific intellectual property, Referee Report, CTAC Ordered list of proposals with grades, etc.

c. Standards and metadata

All the documents will be stored as PDF, word, or excel files.

d. Data sharing

Open Access: Call documents and CTAC reports published on the TA website and statistics generated for EC periodic reports and presentations will be publicly available on ORP website. Membership of CTAC panel (Names, institutions and e-mails of members of the panel) and CTAC general summary of successful proposals (PI, Title, allocations) will also be open access.

Restricted to CTAC members, telescopes operations teams and EC upon request: CTAC results and CTAC Ordered list of proposals (without grades).

Confidential to CTAC members : Proposal forms with personal data and scientific intellectual property, referee reports, technical and user feedback, CTAC Ordered list of proposals with grades are confidential to the CTAC. Possibly shared with EC to show the quality of our feedback. If shared outside of the TAC process personal data needs to be anonymised.

e. Archiving and preservation (including storage and backup)

Open Access documents will be archived and available on the ORP website for the project lifetime and 5 years after the end of the project.

Other data restricted and confidential are stored in the ASTRON servers and are thus available to anyone who has the correct Northstar access rights. The files are stored during the project lifetime and beyond.

V.   Allocation of resources

What are the costs for making data FAIR in your project? How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

There are no immediate costs anticipated and covered by the project to make the datasets produced FAIR. Scientific and CTAC data sets are archived in facility supported archives with a long-term (5 year or longer) perspective. For the technical dataset resulting from the ORP project, they will be deposited in open and free repositories for at least 5 years after the conclusion of the project.

Who will be responsible for data management in your project?

The Executive Management Team, especially the ORP scientific coordinator and ORP project assistant, is responsible for data management within the ORP project, specifically for creating and updating this data management plan (D2) with the support of the responsive leaders.

Facilities are responsible for the data management of scientific data resulting from the Transnational Access programme. On the other hand, PIs are responsible for the data management of scientific publications resulting from the Transnational Access programme, and are encouraged by ORP TA coordinators to apply FAIR and open data sharing practices.

Work Package Leaders (WPL), Task Leads (TL), and Work Package Members (WPM) punctually, are responsible for the data management of technological or software research and development. For technical data and publications resulting from Management and JA 1-2-3-4, it is the burden of the ORP project assistant, the Radio project manager and the OPTICON project manager to support the WPLs with the administrative aspects of outputs and data deposits.

Finally, the data management of the technical and personal data resulting from the CTAC is the responsibility of the CTAC coordinator.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

N/A

VI.   Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)? Is the data safely stored in certified repositories for long term preservation and curation?

The data is secured according to the policies and arrangements of the ORP facilities, which are publicly available (See table 4.1 for the address). They assure a long-term preservation and curation of the data.

VII.   Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing?These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

The ORP project complies with the ethical principles as set out in Article 34 of the Grant Agreement, which states that all activities must be carried out in compliance with Ethical Principles, including the highest standards of research integrity — as set out for instance, in the European Code of Conduct for Research Integrity (European Science Foundation, 2011) — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct.

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

The ORP Project is using the data generated by ORP facilities, which follow their own procedures for the data management. However, since the ORP facilities follow the open policy procedure, no particular influence on the FAIR is expected.

VIII.   Confidentiality

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