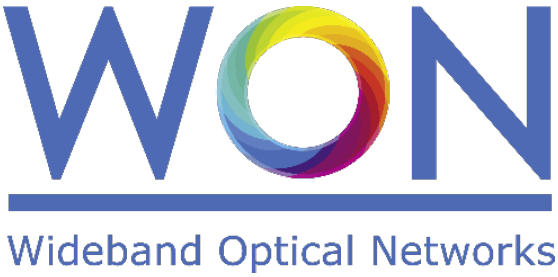
Marie Skłodowska-Curie (MSCA) – Innovative Training Networks (ITN)
H2020-MSCA-ITN European Training Networks



Wideband Optical Networks [WON]

Grant agreement ID: 814276

WP7 Management and governance
Deliverable 7.6 Open Research Data Pilot

DATA MANAGEMENT PLAN

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 814276.

Document Details

Work Package	WP7 Management and governance
Deliverable number	D7.6
Deliverable Title	Open Research Data Pilot. Data management plan
Lead Beneficiary:	Aston University
Deliverable due date:	30 June 2019
Actual delivery date:	29 November 2019
Dissemination level:	Public
Version number	V1.0

Project Details

Project Acronym	WON
Project Title	Wideband Optical Networks
Call Identifier	H2020-MSCA-2018 Innovative Training Networks
Coordinated by	Aston University, UK
Start of the Project	1 January 2019
Project Duration	48 months
WON website:	https://won.astonphotonics.uk/
CORDIS Link	https://cordis.europa.eu/project/rcn/218205/en

Consortium

Beneficiaries	**Partners**
Aston University	Keysight Technologies GmbH
Infinera Germany	Finisar Germany GmbH
Infinera Portugal	Technische Universitaet Berlin
VPIphotonics GmbH	Instituto Superior Technico (IST), University of Lisboa
DTU: Danmarks Tekniske Universitet	Orange
Fraunhofer HHI	
Politecnico di Torino	
Technische Universiteit Eindhoven	
Universiteit Gent	

Abbreviations

CA: Consortium Agreement
DMP: Data Management Plan
DOI: Digital Object Identifier
FAIR: Findable, Accessible, Interoperable and Reusable
EUDAT: Research Data Services, Expertise & Technology Solutions

Contents

1. INTRODUCTION

As part of Horizon2020 the ETN WON project participates in a pilot action on Open Research Data.

This deliverable provides the **WON Data Management Plan (DMP) version 1**, which describes what kind of data will be collected or generated throughout the project and how it will be handled, processed and shared.

Furthermore, the WON DMP will lay out the procedure for data collection, storage, protection, retention of the data. The WON project strives to comply with the open access policy of Horizon 2020 while simultaneously preserving commercial secrets of the participating companies and industries.

Data are collected and processed in accordance with legislative and professional codes of the EU and of countries in which the research is undertaken, as well as the Commission Recommendation of 11 March 2005 on the European Charter for Researchers and the Code of Conduct for the Recruitment of Researchers (www.europa.eu.int/eracareers/europeancharter)

This deliverable is based on the template and the guidelines provided by the European Commission – H2020 templates: Data management plan v1.0 – 13.10.2016.

The DMP is a living document, which will be reviewed, adjusted and updated at the end of each reporting period, or whenever appropriate during the project lifetime (live document).

2. DATA SUMMARY

Being in line with the EU's guidelines regarding the DMP, this document should address for each data set collected, processed and/or generated in the project the following characteristics:

- Dataset description
- Reference and name
- Standards and metadata
- Data sharing
- Archiving and preservation.

1. What is the purpose of the data collection/generation and its relation to the objectives of the project?

Purpose of the data collection and generation. The purpose of data collection is manifold.

First and foremost, data collection is needed to benchmark and reproduce the results obtained within the project. For example, it is important to collect and precisely store all data obtained through experiments and numerical analyses so that it will always be possible to benchmark our newly developed algorithms with respect to a given data set and algorithm. The data will be shared across the ESRs and made publicly available so that researchers outside of WON can compare the performance of their algorithms against ours by using the very same data set.

No less important is to guarantee we use a series of well-defined component characteristics within the project. Some ESRs will work on experiments, some other on numerical analyses. Nevertheless, it is important that we have widely used components characteristics within the Consortium, also within numerical analyses. This will aim at recreating system conditions as close as possible to the reality of the experiments.

Finally, WON is the first project, which aims at considering all aspects of such a wide spectrum. In this context, we will make use of a large set of data of all kinds. The extension of optical transmission bandwidth is seen as one of the options to postpone the upcoming fibre capacity crunch. In this context, it will be of high important we make our data available so that the scientific community will have the possibility to work and compare their discovery against our work. A beneficial side effect of this is that the impact of our related publications will grow significantly.

2. What types and formats of data will the project generate/collect?

In WON two types of data will be generated: for dissemination & reporting and for technical analyses. Within dissemination, we differentiate the following:

1. Deliverables and other EC related documents
2. Management reports
3. Internal scientific reports
4. Publications at conferences (mainly ECOC and OFC) or on highly cited journals

Point 1 - 2 will be carried out by using the following software:

- Microsoft Office 2010 Suite or any other compatible software (such as open source alternatives like LibreOffice and online platform as Google Drive). This will make use of file extensions such as .doc, .docx, .xls, .xlsx, .ppt, .pptx. Also, especially where larger datasets need to be dealt with, .csv and .txt file formats will be used. All finished and approved documents will also be made available as .pdf documents.

- Illustrations and graphic design will make use of Microsoft Visio (Format: .vsd), Photoshop (Format: different types possible, mostly .png), and will be made available as .jpg, .psd, .tiff and .ai files.
- PFDs, PIDs and layouts will preferentially use inkscape.org, an open source software for vector graphics. (Format: .svg), and will be made available as .png, .jpg and .pdf files.

- MP3 or WAV for audio files.

- Quicktime Movie or Windows Media Video for video files. These file formats have been chosen because they are accepted standards and in widespread use. Files will be converted to open file formats where possible for long-term storage.

For what concerns point 3-4, the following software will be used:

- For dissemination, Latex (and the online version ShareLatex) will be the primary choice. Some articles will still be written in Microsoft Word or similar, but the first author will be responsible to convert it into Latex format. This is due to key-target of the consortium to write a book at the end of the project, and Latex guarantees much more flexibility with respect to Microsoft Word

- Block diagram, flow chart and in general, illustration of any kind will be generated by using inkscape, upon agreement with all partners, because a large part of these images will be re-utilized within the book project.

- The data from experiment and numerical analyses will be generated via several software. Not always, it will be possible to save the results in all formats. For instance, some instruments allow only few file extensions, and therefore we need to cope with this. In general, we will guarantee that any important results or dataset will be generated also via open source and shareable formats.

3. Will you re-use any existing data and how?

Yes, some of date is considered to be reused. There are data, for example, concerning the characterization of components, which it makes sense to reuse. Other datasets that will be re-used are the one concerning previous experiments, in particular when these experiments cannot be easily repeated, like field trial with operators. Several partners own these types of datasets and will aim at exploiting them if needed.

4. What is the origin of the data?

The origin of the data, here meant as data for experimental and numerical analysis, will be the instrumentation in the laboratories and the software used to run our analysis (Matlab, Python, etc.).

5. What is the expected size of the data?

This is highly dependent on the context. For what concerns the report and articles, it will be well below 20 MB. For the case of analyses like fiber propagation, it will depend on the file format and accuracy. In general, no large files are planning to be generated, apart from the case of videos.

6. To whom might it be useful ('data utility')?

As mentioned above, the datasets saved from our investigation will be useful for the Consortium, as we will use them to benchmark our algorithms over the project lifetime. In the second moment, they will be useful for the entire scientific community to compare other solutions against those created by the WON Consortium. The articles will be useful for the scientific community as well.

3. FAIR DATA

3.1 Making data findable, including provisions for metadata

1. **Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?**

All data will be given a Digital Object Identifier (DOI) from UK Data Archive.
Data produced at Aston University will be deposited in Aston Research Data Explorer making them identifiable, accessible, discoverable, and reusable.

2. **Will search keywords be provided that optimize possibilities for re-use?**

Keywords associated with research data sets will be linked to the data in the Aston Research Data Explorer. We will identify a common syntax for the files depending on the type of data. These has to be discussed among all partners.

3. **What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.**

- Uploaded metadata will include the following details: data description, authors, school, funders, date of data creation.
- Metadata will be harvested by Google Scholar, re3data.org, EUDAT etc.
- Statistics will be available in terms of number of downloads, number of views, social media clicks etc.

3.2 Making data openly accessible

1. **Which data produced and/or used in the WON will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.**

The WON Consortium partners are taking steps to ensure open access (embargoed where appropriate) to all publications as required by the Horizon 2020 programme. Where available, consortium members will use their institutions' publications repositories to ensure 'green' open access as a minimum. The 'green' route to open access will be followed: the copies of the peer-reviewed publications of the WON Consortium members will be made freely available as soon as possible after publication. Where appropriate, publications will be submitted to the Coordinator's institutional repository (Aston Publications Explorer, https://publications.aston.ac.uk).

Once in the repository and unless embargoed by the publisher or restricted by the consortium, due to ongoing commercial/patent application or re-use of data/release after journal publication, data became available after embargo date. Metadata openly available online by publication date.

When commercial interests prevail and IP protection of research data is necessary, WON consortium members may choose not to make all research data publicly available. Provisions in the WON Consortium Agreement will govern where data will not be shared because of having sensitive or industrial relevance that could be exploited through patenting and commercialization.

2. **What methods or software tools are needed to access the data?**

No specific software tools are required to access the data generated by the WON project.
Data reposted in Aston Data Explorer will be interoperable and be discovered and freely used.

3. **Where will the data and associated metadata, documentation and code be deposited?**

Peer reviewed publications and conference papers will be deposited (embargoed if necessary) in local repositories (such as Aston Data Explorer).The repository services at Aston Data Explorer are free of charge and enable peers to share and preserve research data and other research outputs in any size and format: datasets, images, presentations, publications and software. The digital data and the associated meta-data is preserved through well-established practices such as mirroring and periodic backups. Each uploaded data set is assigned a unique DOI rendering each submission uniquely identifiable and thus traceable and referenceable.

Alternatively, publications and conference papers may also be deposited in the EU open access repository OpenAIRE/Zenodo (https://zenodo.org). Zenodo is a EU-backed portal based on the well-established GIT version control system (https://git-scm.com) and the Digital Object Identifier (DOI) system (http://www.doi.org).

3.3 Making data interoperable

Data produced in WON will be interoperable, i.e. allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. Data will be adhering to standards for formats, be as much as possible compliant with available (open) software applications, and in particular facilitate re-combinations with different data sets from different origins.

1. **What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?**

In order to make data interoperable, data and metadata vocabularies, standards or methodologies will be followed. Data will be given keywords, discoverable names e.g experiment number, date, non-proprietary/open source formats where possible, version numbers, e.g. “Version_1.1”, “Version_1.2” etc, collection methods, date of creation, temporal coverage, authors, funders, ORCIDS, institution, license type and data type.

3.4 Increase data re-use (through clarifying licenses)

1. **How will the data be licensed to permit the widest re-use possible?**

Data will be attributed to Creative Commons license wherever possible.

2. **When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.**

As previously stated, data will be made available immediately upon deposition in the repository unless embargoed by the publisher or restricted by the WON Consortium, due to ongoing commercial/patent application or re-use of data/release after journal publication. Data became available **after embargo date.** Metadata will become openly available online by publication date.

3. **How long is it intended that the data remains re-usable?**

At Aston Data Explorer Data will be stored for a minimum of 10 years.

4. ALLOCATION OF RESOURCES

1. What are the costs for making data FAIR in your project?

Costs related to open access to research data are eligible as part of the Horizon 2020 grant.
In addition, long-term preservation of data in Aston Data Explorer will be carried out by the Aston University, the cost of which is covered by Aston University.

2. Who will be responsible for data management in your project?

The WON project Coordinator with the input from the Project Management team is responsible for data management. The DMP will be revised by the WON Consortium at the end of each reporting period.

3. What resources will you require to deliver your plan?

To implement the DMP, the following resources are required:

- Facilities for storage of large data sets at Aston University and at local repositories of WON partner organisations;
- Submission support from the relevant public data repositories;
- Software to convert the data into standardized form and to provide the required metadata annotation;
- Software to aid the submission of large volumes of data to the repository.

5. DATA SECURITY

1. What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

For the duration of the project and thereafter if appropriate, all research data sets generated by the WON will be stored as outlined below:

Digital data sets	• Will be stores on the responsible partner's local storage systems such as desktop computers, laptops, shared drives. • WPN partner organisations will ensure that the data are stored safely and securely and in full compliance with the European Union data protection laws. • Data will be on secure WON partners' computers and at this stage will only be accessed by members of staff. • After the completion of the project, all the responsibilities concerning data recovery and secure storage will go to the repository storing the data sets. • Some data files that include confidential details will be transferred via secure connections and where appropriate in an encrypted and password-protected form.
Non-digital data	• Will be held in lab books and secure office filing systems. • Paperwork containing identifiable personal data will be shredded immediately once it is no longer needed.

2. Is the data safely stored in certified repositories for long-term preservation and curation?

Data that will be deposited in Aston Data Research Explorer and other WON partners' repositories will be stored and backed up in accordance with local practices.

6. ETHICAL ASPECTS

1. **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).**

Data protection and good research ethics are major topics for the WON Cnsortium. Good research ethics meet all actions to take great care and prevent any situation where sensitive information could get misused. Research data, which contains personal data, will just be disseminated for the purpose for which it was specified by the consortium.
The beneficiaries must process personal data under the GA in accordance with EU and national law on data protection (in particular, Directive 95/46/EC57 and the corresponding national law).
The WON partners have to comply with the ethical principles as set out in Article 34 of the Grant Agreement (Ethics and research integrity), 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 including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct)
- applicable international, EU and national law (in particular, EU Directive 95/46/EC).

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

Personal data is only collected for the purpose of selection and recruitment of personnel to the WON project and is destroyed upon successful appointment. Applicants are made aware that their personal data are shared between all recruiting project partners and have given implicit consent. No other personal data is collected by the WON project.

7. LINKS

1. European Commission data Protection Framework.
   https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

2. European Commission “H2020 Guidelines to the Rule on Open Access to Scientific Publications and open access t0 Research Data in Horizon 2020”.
   https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

3. Zenodo platform.
   https://zenodo.org

4. Digital Object Identifier (DOI) system.
   http://www.doi.org

5. Aston University “Data Protection Policies and Procedure”
   https://www2.aston.ac.uk/data-protection

6. Aston University Data Repository “Aston Data Explorer”
   http://researchdata.aston.ac.uk/

7. EUDAT – Research Data Services, Expertise & Technology Solutions
   https://eudat.eu/