Data Management Plan

WOSCAP
Whole of Society for
Conflict Prevention and Peacebuilding

Deliverable 1.3

(Version 1; 30 July 2015)

Prof. dr. ir. Georg Frerks & Toon Dirkx MA

Horizon 2020 | BES-12-2014/2015

Table of Contents

1. Introduction

Research data form the basis of the WOSCAP project. They play a crucial role and should be effectively managed to ensure the verification and reuse of research results, and the sustainable storage of the dataset.[1]

This Data Management plan aims at providing a timely insight into facilities and expertise necessary for data management both during and after the WOSCAP research, to be used by all WOSCAP researchers and their environment: including WOSCAP's Executive Board, Steering Group, work package (WP), thematic and case study leaders, research funders, research users and data supporters.

The most important reasons for setting up this Data Management plan are:
- Embedding the WOSCAP project in the EU policy on data management, which is increasingly geared towards providing open access to data that is gathered with funds from the EU. The rationale is that the Horizon 2020 grant consists of public money and therefore the data should be accessible to other researchers;[2]
- Enabling verification of the research results of the WOSCAP project;[3]
- Stimulating the reuse of WOSCAP data by other researchers;[4]
- Enabling the sustainable and secure storage of WOSCAP data in the DataverseNL and DANS EASY repositories;[5]
- Helping to streamline the research process from start to finish. The data management plan clarifies in advance the required data expertise and facilities to store data.[6]

This document contains the main elements of a research data management plan and is based on various Dutch and European sources. The notion of "data" in the WOSCAP project can be understood in a very broad sense. The DANS (2015) brochure "Data management plan for scientific research" gives more detailed information.

The Data Management plan is a document that is submitted to the EU as project **deliverable 1.3** in July 2015. It is important to note however that the document will evolve and further develop during the project's life cycle. It can be identified by a version number and a date. Updated versions will be uploaded by Utrecht University (UU), which is the primary responsible for data management. WOSCAP partners can forward questions and suggestions, as to (additions to) the contents and use

---

[1] DANS. (2015). "*Research Data Management Plan for Scientific Research.*"
http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material/DANSdatamanagementplanmei2015UK.pdf [Accessed on 8 June 2015].
[2] European IPR Helpdesk. (2014). *Fact Sheet Open Access to publications and data in Horizon 2020: Frequently Asked Questions (FAQ)*.
[3] DANS. (2015). "*Research Data Management Plan for Scientific Research.*"
http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material/DANSdatamanagementplanmei2015UK.pdf [Accessed on 8 June 2015].
[4] Ibid.
[5] Ibid.
[6] Ibid.

of the data management plan to UU, and will be informed when a new version will be uploaded in the online workspace of WOSCAP and Dataverse.

Two repositories will be used to store data. During the project's duration data will be stored and managed in **Dataverse**.[7] At the completion of the project, when all data is gathered, UU will transfer the final dataset to **DANS EASY**.[8]

2. Data Set Description

The WOSCAP project builds on human security oriented and participative, dialogic approaches to research and analysis. Much policy relevant research is done on the basis of an analysis of policy documents and interviews of involved policy makers and the respective organisations' staff. Though this project will also include this, it moves beyond this approach by involving a larger range of stakeholders and beneficiaries. Project partners have pioneered a number of innovative tools to be consistent with a more bottom-up approach than classic approaches to data gathering and analysis. These emphasise a role for qualitative methods, used in conjunction with appropriate quantitative data to provide more detailed insights into local experiences and peacebuilding processes.[9]

This data, or in more academic terms 'empirical evidence of social life'[10], roughly comes in three forms: talk, text and observable modes of interaction. In order to gather these different forms of data, various data collection techniques will be employed. Qualitative data collection techniques that can be used are for example (in-depth) interviews, focus group discussions and observations. A quantitative data collection technique that may be used is a survey questionnaire. The project will also make use of secondary sources, including: literature research, policy documents, archives and databases. Altogether, this mixed-method approach will lead to a convergence of multiple sources of evidence.[11]

The data for the WOSCAP project will be gathered by various researchers. Primary data will mainly be gathered in the case study countries (Ukraine, Yemen, Georgia and Mali), while secondary sources may be gathered anywhere through the use of online accessible information.

3. Data Storage

During the life cycle of the project data will be stored and systematically organised in a repository called 'Dataverse'.

---

[7] See: https://dataverse.nl/dvn/ and http://dataverse.org/

[8] See: https://easy.dans.knaw.nl/ui/home

[9] This section is derived from the WOSCAP DoA (as in the GA), p. 12.

[10] See: Ragin, C. (1994). *Constructing Social Research: the Unity and Diversity of Method.* Thousand Oaks: Pine Forge Press.

[11] See: Yin, R. K. (2003). *Case Study Research: Design and Methods.* Thousand Oaks: Sage Publications, Inc.

3.1 Dataverse

Dataverse is an open source web application to share, preserve, cite, explore and analyse research data. It facilitates making data available to others, and allows to replicate others work. A Dataverse repository hosts multiple dataverses. Each dataverse contains datasets or other dataverses, and each dataset contains descriptive metadata and data files (including documentation and code that accompany the data).[12]

Dataverse was developed at Harvard University in 2006. Nowadays, there are several Dataverse repositories installed in universities around the world. The Dutch Dataverse Network was initially created by staff members at the library of Utrecht University to provide their local research community with a solution to store a wide variety of scientific data (texts and raw research data, but also video material and complete databases) in a safe and sustainable way.[13]

Figure 1 gives a schematic overview of the global Dataverse Network.

Figure 1 – Dataverse Network[14]

---
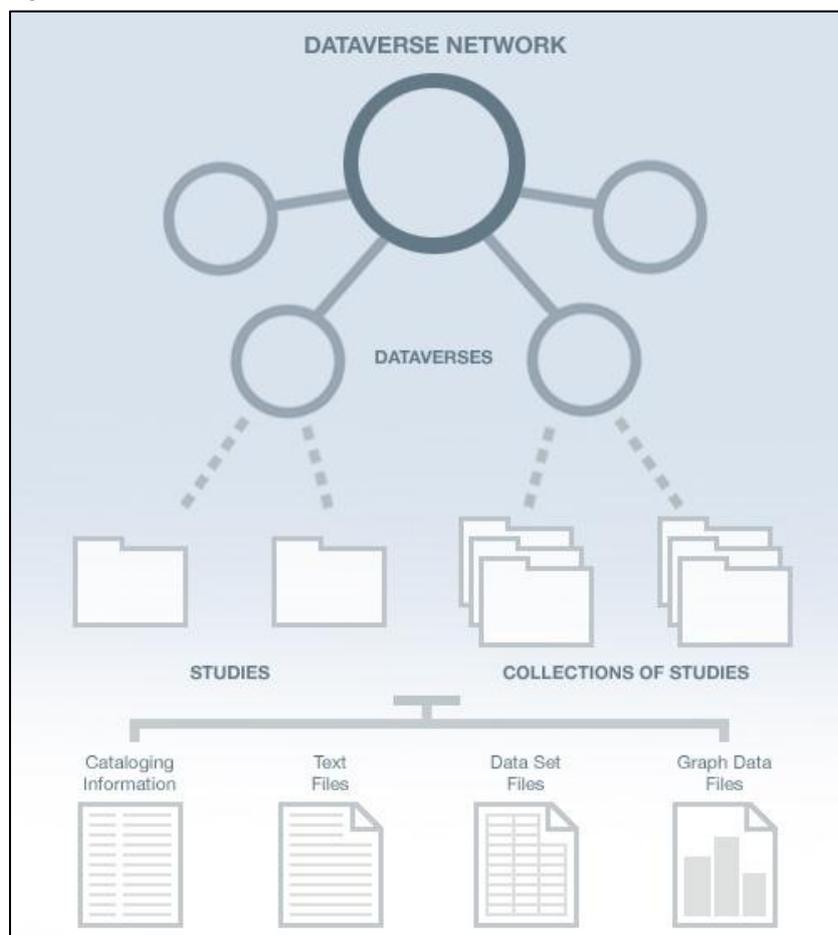
[12] See: http://dataverse.org/about/ [Accessed on 11 June 2015].

[13] See: Van Dijk, P. (2014). The Dutch Dataverse Network, an inter-institutional collaboration effort facilitated by SURFconext.
https://blog.surfnet.nl/dutch-dataverse-network-inter-institutional-collaboration-effort-facilitated-surfconext/ [Accessed on 11 June 2015].

[14] Figure taken from http://www.dlib.org/ [Accessed on 12 June 2015].

For the WOSCAP project, UU has created a 'WOSCAP Dataverse' in the Dutch Dataverse Network. During the life cycle of the project WOSCAP researchers will get access to the WOSCAP Dataverse, which is managed by UU. **Figure 2** shows how the WOSCAP Dataverse is envisioned to be structured.

Figure 2 – Possible structure of WOSCAP Dataverse[15]
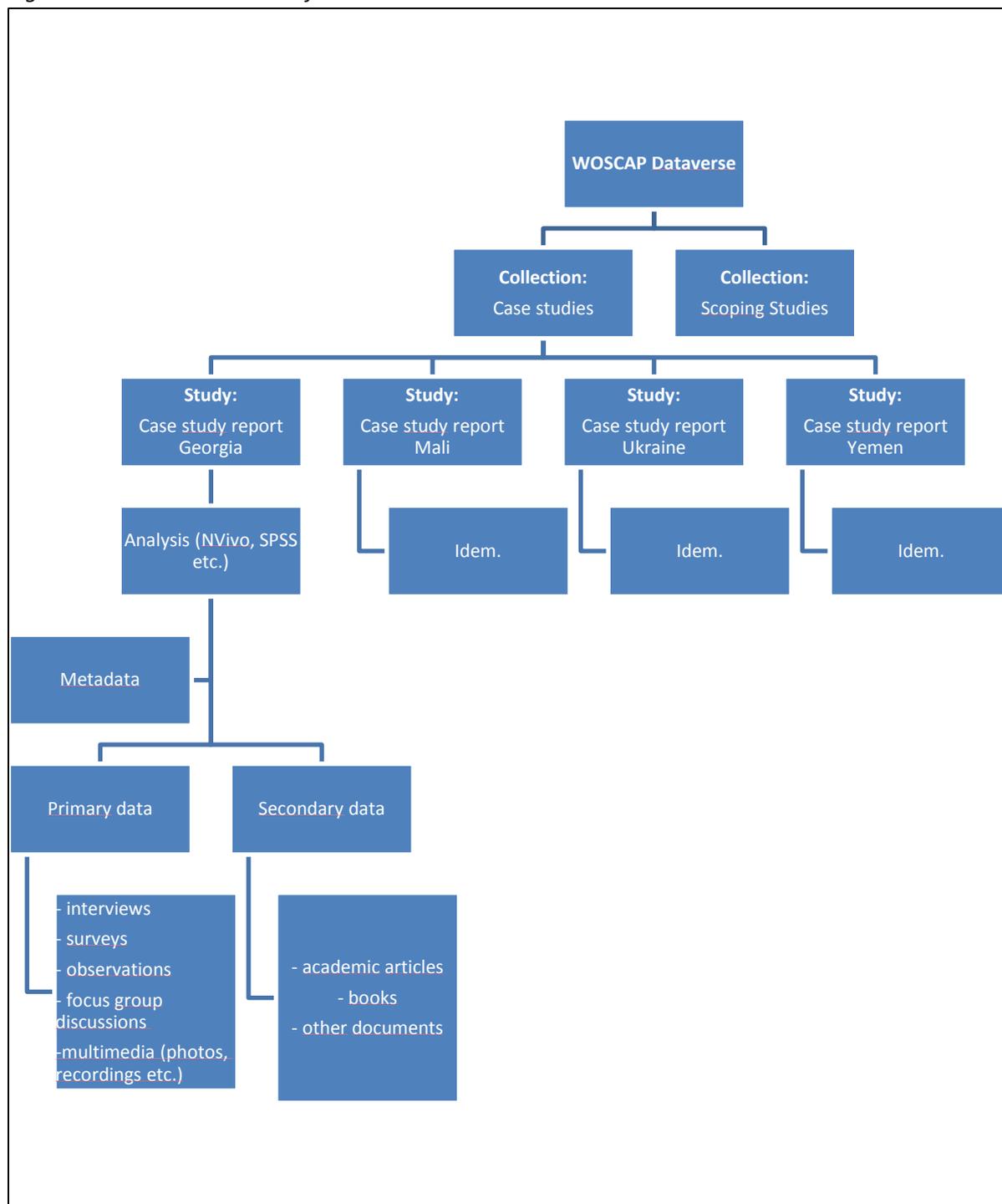
3.1.1 Roles in Dataverse

In the WOSCAP Dataverse different roles are distinguished in accordance with the responsibilities of the involved partner organisations. **Table 1** shows what these roles entail.

Table 1 – Roles in Dataverse[16]

	Administrator (UU)	Curator (UU)	Contributor (WP leaders & Researchers)	User
Browse collections and search for studies	✔	✔	✔	✔
Access and Analyse Data	✔	✔	✔	✔
Create Study and Upload Files	✔	✔	✔	
Set Permission to Study	✔	✔	✔	
Update Study After Release (make draft)	✔	✔		
Release Study	✔	✔		
Organise Study by Collection	✔	✔		
Add new admins, curators, and contributor	✔			
Change Settings (contact email, DV name, etc)	✔			
Release Dataverse when ready	✔			

---

[16] Table 1 is derived from Pronk, T. (2015). *Dutch Dataverse Network*. Presentation.

3.1.2 Process

The process of uploading data to the WOSCAP Dataverse contains several steps. **Figure 3** gives a schematic overview of these steps:

Figure 3 – Process WOSCAP Dataverse[17]
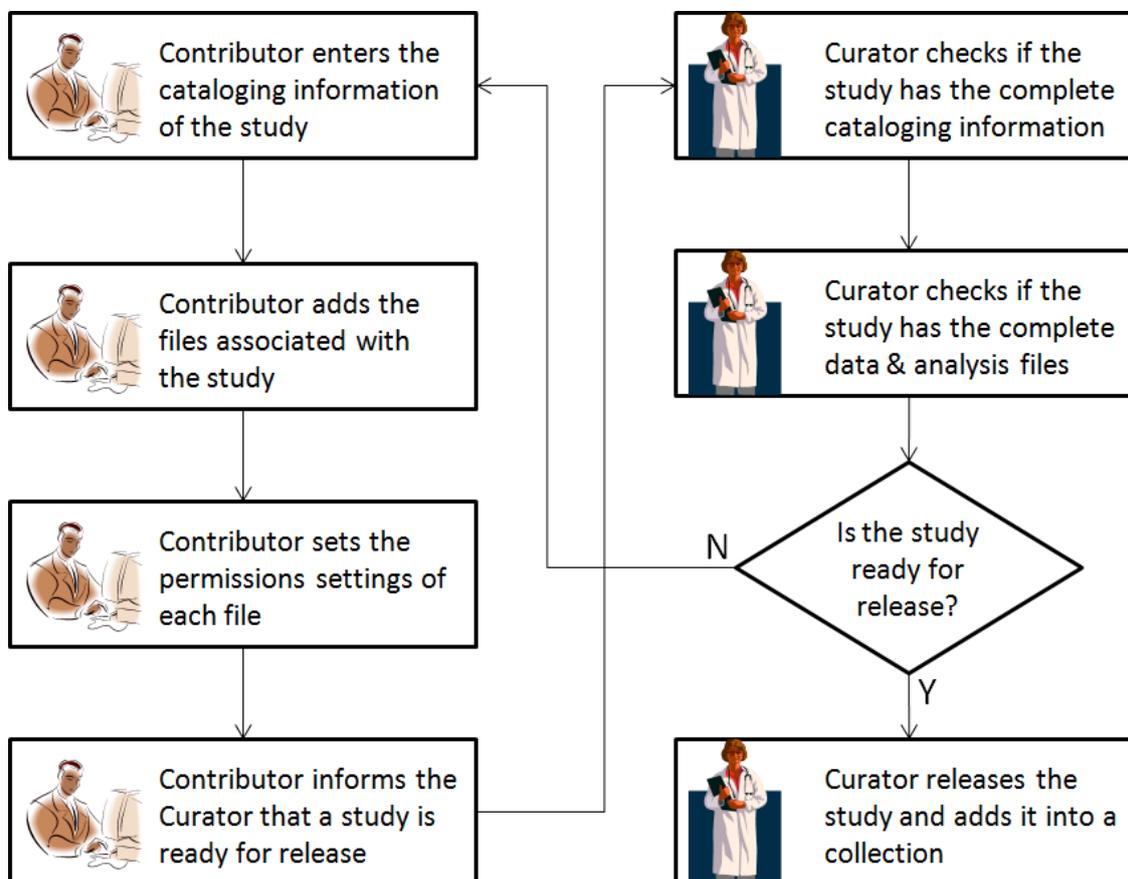
3.1.3 Access to the dataset in Dataverse

The Dataverse Network encourages the open access of research data.[18] It is possible however, to restrict access to those cases in which the author wants to limit the use or access of the data. In the WOSCAP project this may occur when the security of researchers and respondents is in danger or when problems arise with interpreting the contextual specificity of the data by other researchers. Therefore, there are three levels of access control:

- **A public study with terms of use**: in this case the descriptive information for the study can be viewed without any conditions, but the data files can only be accessed after the user agrees to the terms of use assigned to that study.[19]

---

[17] Figure 3 is derived from Pronk, T. (2015). *Dutch Dataverse Network*. Presentation.
[18] This sub-paragraph is based on: Crosas, M. (2011). The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data. *D-Lib Magazine*, 17(1). See also http://dataverse.org/
[19] Crosas (2011).

- **A study with restricted file(s)**: in this case the descriptive information remains public, but one or more data files are restricted. Only password-authorized users are allowed to view and download the restricted files.[20]
- **A restricted study**: when the entire study is restricted, the metadata are still searchable, making part of the description discoverable, but access is not allowed to either the full cataloguing information or the data files.[21]

External users may request access to a restricted file or a full study, which can subsequently be granted by the Dataverse owner. Authorisation to a set of studies can also be given based on a range of IP addresses associated with a university or other institution. Ultimately, however, WOSCAP data authors have control over who can access their data.

3.1.4 Citation of the dataset

The WOSCAP Dataverse will receive a unique citation in order to ensure that other researchers, who use the dataset, include it as a reference to the original dataset of the WOSCAP project.[22]

3.2 Metadata

Metadata is data on the research data themselves. It enables other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers, can better determine whether the dataset is relevant and useful for their own research.[23]

In the WOSCAP Dataverse, metadata (type of data, location, population etc.) will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data.

4. Data Security

The WOSCAP project will use methods that emphasise good field access and extended contact and trust building with participants, and they are also sensitive to ethical concerns in doing social science research among respondents in conflict-affected areas. Due to the sensitive nature of some of the topics that will be discussed in, for example interviews in the case study countries, data security is of vital importance.[24]

The following guidelines will be used in order to ensure the security of the data:
- Keep anonymised data and personal data of respondents separate;
- Encrypt data if it is deemed necessary by the local researchers;
- Store data in at least two separate locations to avoid loss of data;

---

[20] Crosas (2011).

[21] Ibid.

[22] Pronk, T. (2015). *Dutch Dataverse Network*. Presentation.

[23] DANS. (2015). *"Research Data Management Plan for Scientific Research."*
http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material/DANSdatamanagementplanmei2015UK.pdf [Accessed on 8 June 2015].

[24] See also the 'Ethics and Societal Impact' paragraph of the WOSCAP DoA (as in the GA), pp. 66-71.

- Limit the use of USB flash drives;
- Save digital files in one the preferred formats (see Annex 1), and;
- Label files in a systematically structured way in order to ensure the coherence of the final dataset (see Annex 2).

5. Data Sharing

The nature and location of the research subject inevitably inhere sensitive and controversial topics related to armed violent conflict. Topics raised may be linked to deeply felt socio-cultural, ethnic, religious, linguistic or regional identities that are linked to feelings of grievance and deprivation on the one hand, and to suffering, bereavement and grief on the other. Discussing both war and peace will touch upon those issues and is unavoidable to collect the data needed to do this research. Therefore, utmost care is taken in this project to do this with the required precautions and highest levels of responsibility.[25]

These precautions and other measures such as the privacy, anonymity and informed consent procedures, the code of conduct, and the safety of participants and researchers, ethics are specified in the WOSCAP DoA (as in the GA, pp. 66-71). It is important to keep this section of the DoA in mind, while reading the following two paragraphs on data sharing.

5.1 Public Availability of Data

In view of the precautions for protection of personal data, it is explicitly confirmed that the data collected will be publicly available, after care is taken with regard to rules of confidentiality, anonymity and protection. Anonymised final data sets will be open access and procedures are set as to how data will be preserved and archived after closure of the programme in the DANS EASY repository. We are aware of post-publication risks to local researchers and end-users in our research sites and will mitigate all reasonable risk before publication. We will not use any images or names of any research subjects or local researchers without their prior consent and agreement.[26]

5.2 Opt Out

It is important to note that even though comprehensive measures are taken to ensure the safety of participants, researchers and their environment, it is only after a WOSCAP report or scholarly article is published, that the question of open access arises. Open access does not entail an absolute obligation to publish all data[27]: it is up to researchers and UU to decide whether data is suitable and ethical to be published or not.

---

[25] This section is derived from the WOSCAP DoA (as in the GA), p. 69.

[26] Ibid.

[27] European IPR Helpdesk. (2014). *Fact Sheet Open Access to publications and data in Horizon 2020: Frequently Asked Questions (FAQ)*.

6. Preservation of the Final Dataset

Dataverse is a useful tool to store the data during the life cycle of the project, but it is not geared towards the long-term preservation of the data. Therefore, at the completion of the project, the final dataset will be transferred to the DANS EASY repository, which ensures sustainable archiving of the final research data.[28]

This DANS EASY repository is a service offered by DANS (Data Archiving and Networked Services), which promotes sustained access to digital research data.[29] DANS is an institute of KNAW[30] (The Royal Netherlands Academy of Arts and Sciences) and NWO[31] (The Netherlands Organisation for Scientific Research).

DANS EASY is certified by the Data Seal of Approval and provides the WOSCAP data set with a 'persistent identifier'[32] (a long-lasting reference to a digital object, in this case the WOSCAP dataset). UU will manage this process of transferring the final dataset from Dataverse to DANS EASY. The use of DANS EASY by the WOSCAP project is free of charge, since these expenses are covered at the institutional level by Utrecht University.

---

[28] See: https://easy.dans.knaw.nl/ui/home

[29] See: http://www.dans.knaw.nl/en

[30] See: https://www.knaw.nl/en?set_language=en

[31] See: http://www.nwo.nl/en

[32] See: https://easy.dans.knaw.nl/ui/home

References

Crosas, M. (2011). The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data. *D-Lib Magazine*, 17(1).

Guéret, C. (2013). *How to publish Open Data on the Web.* In: Pilot Linked Open Data, (pp. 115-120).

DANS. (2015). *DANS Preferred Formats*. Retrieved 8 June, 2015, from
http://www.dans.knaw.nl/en/deposit/information-about-depositing-data/

-------(2015). *Research Data Management Plan for Scientific Research*. Retrieved 8 June, 2015, from
http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material

-------(2015). *Template for Research Data Management Plan*. Retrieved 8 June, 2015, from
http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material

-------(2015). *Your 7 steps to sustainable data*. Retrieved 1 June, 2015, from
http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material

European Commission. (2013). *Guidelines on Data Management in Horizon 2020*. Retrieved 2 June, 2015, from
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European IPR Helpdesk. (2014). *Fact Sheet Open Access to publications and data in Horizon 2020: Frequently Asked Questions (FAQ)*. Retrieved 3 July, 2015, from
https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf

Pronk, T. (2015). *Dutch Dataverse Network*. Presentation.

Ragin, C. (1994). *Constructing Social Research: the Unity and Diversity of Method.* Thousand Oaks: Pine Forge Press.

Van Dijk, P. (2014). The Dutch Dataverse Network, an inter-institutional collaboration effort facilitated by SURFconext. Retrieved 16 June, 2015, from
https://blog.surfnet.nl/dutch-dataverse-network-inter-institutional-collaboration-effort-facilitated-surfconext/

Annex 1 – Preferred Formats[33]

1.1 Selection of File Formats

All formats of digital files stand the risk of becoming obsolete in the future. If a file format becomes obsolete, it means that the current software will not be able to represent and use the content of the file in the way it was meant to at the time of creation.

To prevent file format obsolescence, some precautions can be taken. One such measure is to select file formats which have a high chance of remaining usable in the far future. As a general guideline, DANS considers that the file formats best suited for long-time preservation and accessibility:

1. are commonly used;
2. have open specifications;
3. are independent of specific software, developers or suppliers.

However, it is not always possible to select formats that meet with all of these ideal attributes.

1.2 Preferred and Acceptable Formats

DANS has assessed a number of file formats resulting in a list of preferred formats and acceptable formats. This list will change over time as new formats will be developed and others will fall into disuse.[34]

Preferred formats are the file formats which can be trusted to offer the best long-term guarantees for usability, accessibility and robustness. In principle, DANS expects these formats to be durable for the long term. DANS will accept research data deposited in preferred formats in DANS EASY without question.

Acceptable formats are file formats which are commonly used besides the preferred formats; have average to reasonable scores regarding their usability, accessibility and robustness in the long term. DANS strongly prefers the use of preferred formats but in most cases, the use of acceptable formats will be allowed in to the archive as well.

UU therefore strongly recommends WOSCAP data depositors only to deliver their data in the preferred format corresponding to the type of data. Please contact UU if your data is stored in file formats other than the preferred formats listed below.

Table 1 presents a summarised overview of the DANS Preferred and Acceptable Formats.

Table 1 – Preferred and acceptable formats for data storage

	Preferred format(s)	**Acceptable format(s)**
Text documents	• PDF/A (.pdf)	• OpenDocument Text (.odt) • MS Word (.doc, .docx)

---

[33] This annex is based entirely on the DANS document on 'Preferred Formats'. During a meeting with DANS on 10 June 2015 DANS confirmed their document on preferred formats may be used to inform the WOSCAP Data Management Plan.

[34] May this happen during the lifecycle of the project, UU will update WOSCAP partners on the preferred and acceptable formats.

		• Rich Text File (.rtf)
		• PDF (.pdf)
Text file	• Unicode TXT (.txt, …)	• Non-Unicode TXT (.txt, …)
Marked-up language		• XML (.xml)
		• HTML (.html)
Spreadsheets	• OpenDocument Spreadsheet (.ods)	• MS Excel (.xls, .xlsx)
	• Comma Separated Values (.csv)	• PDF/A (.pdf)
		• OOXML (.docx, .docm)
Databases	• ANSI SQL (.sql, …)	• MS Access (.mdb, .accdb)
	• Comma Separated Values (.csv)	• dBase III or IV (.dbf)
Statistical data	• R	
	• SPSS Portable (.por)	
	• SAS transport (.sas)	
	• STATA (.dta)	
Images (raster)	• JPEG (.jpg, .jpeg)	• JPEG 2000 (.jp2)
	• TIFF (.tif, .tiff)	
	• PNG (.png)	
Images (vector)	• Scalable Vector Graphics (.svg)	• Adobe Illustrator (.ai)
		• PostScript (.eps)
Video	• MPEG-2 (.mpg, .mpeg, …)	
	• MPEG-4 H264 (.mp4)	
	• Lossless AVI (.avi)	
	• QuickTime (.mov)	
Audio	• WAVE (.wav)	• MP3 AAC (.mp3)
Computer Aided Design (CAD)	• AutoCAD DXF versie R12 (.dxf)	• AutoCAD other versions (.dwg, .dxf)
Geographic Information (GIS)	• Geographic Markup Language (.gml)	• ESRI Shapefiles (.shp and associated files)
	• MapInfo Interchange Fomat (.mif/.mid)	• MapInfo (.tab and associated files)
		• Keyhole Markup Language (.kml)
Images (georeferenced)	• GeoTIFF (.tif, .tiff)	• TIFF World File (.tfw en .tif)
Raster GIS	• ASCII GRID (.asc, .txt)	• ESRI GRID (.grd and associated files)
3D	• WaveFront Object (.obj)	• COLLADA (.dae)
	• X3D (.x3d)	• Autodesk FBX (.fbx)
RDF	• W3C standards	

Annex 2 – Labelling Files

Labelling files in a systematically structured way is vital to ensure the coherence of the final dataset of the WOSCAP project. Therefore, this annex provides frameworks to label primary research data and a way to manage different versions of project deliverables.

2.1 Labelling Primary Research Data

Table 2 provides a framework for the labelling of primary data.

Table 2 – Labelling primary data

Type of data(*)	Sub project(*)	Country name	Topic(*)	Deliverable	Organisation	Number
INT (Interview) FGD (Focus group discussion) OBS (Observation) SUR (Survey)	CS (Case study)	UA (Ukraine) YE (Yemen) GE (Georgia) ML (Mali)	GOV (Governance) SSR (Security Sector Reform) CP (Conflict Prevention)	D1.1 - D6.8 See list of deliverables	- GPPAC - UU - LSE - IRENE - BF - ECP - IWP - PDF - TSU - USJPB	01, 02, 03 etc.

* The list of topics will be expanded during the life cycle of the project

Based on the above table, possible file names could be:
1. INT-CS-UA-GOV-D3.4-IWP-01
2. SUR-CS-GE-SSR-D3.2-TSU-14
3. FGD-CS-YE-CP-D3.5-PDF-08

The first file name is the label of an interview for the case study in Ukraine on the topic of governance, part of deliverable 3.4, conducted by the Institute of World Policy with interview number 01.

The second file name is the label of a survey for the case study in Georgia on the topic of Security Sector Reform, part of deliverable 3.2, and conducted by Tbilisi State University with survey number 14.

The third file name is the label of a focus group discussion for the case study in Yemen on the topic of conflict prevention, part of deliverable 3.5, conducted by the Political Development Forum with focus group discussion number 08.

2.2 Managing Different Versions of Deliverables

The WOSCAP project will also use a systematic structure to ensure the effective labelling of digital project deliverables and the management of different versions of these deliverables. **Table 3**, which is inspired by the list of deliverables in the Grant Agreement, provides a framework for labelling these project deliverables.

Table 3 – Labelling Deliverables

Deliverable	Deliverable Name	WP	Lead	Delivery Date	Version (Date)
D1.1 – D6.8	PME Framework, Quality Assurance Plan, etc.	1 – 6	- GPPAC - UU - LSE - IRENE - BF - ECP - IWP - PDF - TSU - USJPB	M1 – M29	e.g. 31 July 2015

Based on the above table, possible file names could be:
1. D2.2-Scoping study Gender-2-ECP-M6-01 November 2015
2. D4.6-Review of EU policy institutional level-4-IRENE-M18-14 November 2015
3. D1.3-Data Management Plan-1-UU-M2-27 July 2015

The first file name is the label of deliverable 2.2 which is a scoping study on gender, part of work package 2, written by the Escola Cultura de Pau, the delivery date is M6 and the version is of 01 November 2015.

The second file name is the label of deliverable 4.6 which is a review of EU policy in the institutional level, part of work package 4, written by the Institute for Research and Education on Negotiation, the delivery date is month 18 and the version is of 14 November 2016.

The third file name is the label of this report, as it is submitted to the European Commission.

Annex 3 – Important Links

Dataverse NL – Dutch Dataverse Network
https://dataverse.nl/dvn/
* By clicking on this link WOSCAP WP leaders and researchers can access the WOSCAP Dataverse. UU will send instructions to the WOSCAP partners on how to acquire a username and password.

Dataverse – Global Dataverse Network
URL: http://dataverse.org/

DANS – Data Archiving and Networked Services
http://www.dans.knaw.nl/en

DANS EASY – DANS Research Data Repository
https://easy.dans.knaw.nl/ui/home
* The final dataset of the WOSCAP project will be transferred by UU to this repository where it will get a persistent identifier.

KNAW – The Royal Netherlands Academy of Arts and Sciences
https://www.knaw.nl/en?set_language=en

NWO - The Netherlands Organisation for Scientific Research
http://www.nwo.nl/en