D8.1 Data Management Plan

Project number: 722581

Project acronym: PLATO

Project title: The Post-crisis Legitimacy of the European Union

Due date of deliverable: June 2017 (Month 6)

Actual delivery date: July 2017 (Month 7)

Lead beneficiary of this deliverable: 1 – ARENA (UNIVERSITETET I OSLO)

Dissemination level: Confidential, only for members of the consortium (including the Commission Services)

Document revision history

Version	Date	Modification(s)	Author(s)
V0.1	12/06/2017	Document creation	Marit Eldholm
V0.2	15/06/2017	First draft	Marit Eldholm
V0.3	06/07/2017	Amended after internal review (members of PLATO Steering Committee 2 and UiO central support unit staff member responsible for open science)	Marit Eldholm
V1.0	07/07/2017	First version submitted to EC	Marit Eldholm
V2.0	31/01/2019	1st planned revision	
V3.0	31/01/2021	Final version	

List of terms / abbreviations

Term	Definition
CC	Creative Commons, organisation releasing copyright licences free of charge
DANS	Data Archiving and Networked Services, The Netherlands Institute for Permanent Access to Digital Research Resources
Dataset	A collection of data
Dataset description	The metadata about a dataset
DDI	Data Documentation Initiative (international standard for describing data from the social sciences)
DMP	Data Management Plan
ESR	Early Stage Researcher, PhD candidate recruited by PLATO
Metadata	A number of defined characteristics or attributes of data
NSD	Norwegian Centre for Research Data
PID	Persistent identifier, a long-lasting reference to a digital file
PLATO	The Post-Crisis Legitimacy of the European Union (EU-funded project)

Table of contents

Executive summary

This is the first version of the PLATO Data Management Plan (DMP). The DMP ensures that the data handling within the PLATO project is both legally and ethically correct and methodologically sound. It describes the types of data that will be generated and collected during the project, the standards that will be used, how the data will be exploited and shared for verification and reuse, as well as the preservation and storage of data. In its next version, the application of the DMP to 15 individual PhD projects will be further detailed.

1. Introduction

1.1. Background

PLATO participates in the Horizon 2020 Pilot on Open Research Data. PLATO's Data Management Plan (DMP) describes the types of data that will be generated and collected during the project, the standards that will be used, how the data will be exploited and shared for verification and reuse, as well as the preservation and storage of data.

The provisions for open access to research data are described in Grant Agreement Arts 29.3 *Open access to research data* and 39.2 *Processing of personal data by the beneficiaries*. This document follows the template for the H2020 Data Management Plan v1.0 (13/10/2016).

1.2. Objectives of the PLATO project

PLATO investigates the legitimacy of the EU's responses to the financial crisis. It uses the example of the financial crisis to build and test theory of what would amount to a legitimacy crisis in the case of a multi-state, non-state political system such as the EU. PLATO generates new understandings of the EU's legitimacy crisis through multiple, connected case studies undertaken by 15 Early Stage Researchers (ESRs).

Six projects investigate different actors with whom the EU needs to be legitimate. To what extent have the EU's crisis responses been contested by different actors? Actors to be studied are EU member states, sub-national authorities, EU institutions, agencies, networks and courts, social partners/lobby groups, and actors exposed to policy failure.

Nine projects investigate whether the EU's crisis responses meet standards of democratic legitimacy, including parliamentary representation, anti-corruption, non-domination, political trust, European identity, civil society, political competition, public sphere, and elections.

1.3. Project timeline and DMP versions

The 15 ESRs employed in PLATO will start their work on 01/09/2017 or 01/10/2017. Table 1, which outlines the types of data to be generated and collected within each project, is based on the project proposals submitted at the recruitment stage and are as such preliminary only. The individual PhD projects will be presented and discussed at the PLATO kick-off conference on 18-19 October 2017, and further developed in cooperation with each individual supervision team. The foreseen research design, including methods and data, is therefore subject to revision. The individual collection of data is expected to be finished in time for the next revision of this document. The DMP will be revised after the end of the first reporting period (31/12/2018) and again at the end of the project. The subsequent versions of this DMP

are foreseen by 31/01/2019 (v2.0) and 31/01/2021 (v3.0) and will include information on all datasets created by PLATO.

1.4. Types of data generated and collected

PLATO's ESRs will conduct social science research that involves: (1) generating raw data in the form of semi-structured interviews, surveys and focus groups; and (2) collecting publicly available documents for textual/discourse analysis, such as official documents, parliamentary debates, media debates, surveys, legal decisions, and voting patterns.

For the data that is already available in the public domain (2), references to the original data will be made in the publication where the data is used in the form of overviews/list of analysed files. The raw files as such will not be archived for reasons of ownership and copyright.

Table 1: Types of data generated/collected in PLATO

Data	Description	File format	ESR
Generated			
Semi-structured interview	Open interview based on an interview guide with an informal grouping of topics and questions to be adapted to the interview setting, allowing new ideas to be brought up during the interview as a result of what the interviewee says. Several ESRs will conduct semi-structured interviews with key actors in specific policy processes, such as decision makers, member state representatives, EU institutions.	• Text (notes, transcripts): **PDF (.pdf)** • Audio: **WAVE/BWF (.wav)** (MP3 .mp3)	ESR1 ESR4 ESR5 ESR6 ESR8 ESR12 ESR13
Focus group	Interviews with a demographically diverse group of people; guided or open discussions about perceptions, opinions, beliefs, or attitudes towards i.e. specific EU policies. Participants can learn from each other as they exchange views. The researcher either takes notes or records the discussions. Random or purposive sampling of individual units, using online questionnaires to identify respondents.	• Text (notes, transcripts): **PDF (.pdf)** • Audio: **WAVE/BWF (.wav)** (MP3 .mp3) • Video: **QuickTime (.mov), MPEG-4 H.264 (.mp4), LosslessAVI (.avi)**	ESR9 ESR11
Survey	Questionnaires with standardised answers; either measuring separate variables (party preference, behaviour, facts) or aggregated into scales (attitudes). Sampling of individual units from a population. Data collection via online surveys. Cross-sectional study or successive independent sample studies.	• Text (notes, transcripts): **PDF (.pdf)** • Statistical data: **SPSS (.sav), SPSS Portable (.por), STATA (.dta), DDI (.xml), data (.csv) + setup (.txt)**	ESR2 ESR12

Collected	Source		
Official documents	Reports (EU, GRECO, other international sources), laws, historical documents, journals, academic papers, EU treaties, legislative databases (Prelex), meeting protocols, intergovernmental bargaining (Council of Ministers)	• Text documents: **PDF (.pdf)** • Spreadsheets: **Excel (.xls, .xlsx)**	ESR5 ESR6 ESR8 ESR10 ESR12
Parliamentary debates	Transcripts of national parliamentary and European Parliament debates publicly available from the respective parliament's website	• Text documents: **PDF (.pdf)**	ESR7 ESR13
Media	Media debates: a selection of sources from selected countries. News reports, analyses and commentaries on a certain decision, policy, institution etc. A selection/combination of online and print news media, television and radio. In most cases ESRs can access newspapers through university library subscriptions. Where that is not possible, single subscription fees will be paid.	• Text documents: **PDF (.pdf)**	ESR1 ESR7 ESR11 ESR13 ESR14 ESR15
ECJ decisions	ECJ decisions and articles/commentaries on these decisions by national and European judges published in specialised reviews	• Text documents: **PDF (.pdf)**	ESR4 ESR10
Personal CVs	Network analysis: CVs of European and national judges and lawyers to uncover partnerships, academic and/or professional allies	• Text documents: **PDF (.pdf)**	ESR4
Surveys	Eurobarometer surveys European Election Survey Various international surveys on corruption	• Statistical data: **SPSS (.sav), SPSS Portable (.por), STATA (.dta), DDI (.xml), data (.csv) + setup (.txt)**	ESR8 ESR10 ESR15
European election results	Quantitative analysis of voting patterns	• Statistical data, text or spreadsheets	ESR11

1.5. Data utility

The data generated and collected by PLATO might be useful for the broader scientific community for the sake of validating results presented in PLATO's scientific publications, as well as for future social science research.

2. FAIR data

PLATO complies with the guiding principles for scientific data management outlined by the H2020 programme, which ensure that the research data is **Findable**, **Accessible**, **Interoperable** and **Reusable** (FAIR).

2.1. Making data findable (F)

Persistent identifiers (PID)

Upon archiving in a data repository, each dataset will be assigned a persistent identifier, a unique code in the form of an URL or DOI. As a general rule, PLATO's datasets will be archived in the DANS data archive, the Netherlands Institute for Permanent Access to Digital Research Resources, which assigns and use DOIs as a PID.

Metadata

Each dataset produced will be described with metadata to allow for discovery and analysis and to support effective data sharing. As a general rule, PLATO will use the Data Documentation Initiative (DDI),[1] which is an international standard for describing data from the social sciences, including surveys, questionnaires, and statistical data files.

The metadata for each dataset will contain information on the following (depending on the nature of the data): creator, access conditions, collection methods, time references, structure and organisation of data files, file formats, variable names, labels and descriptions of variables and values, codes for missing values, and hard- and software used to process and analyse the data.

Thematic search keywords will be provided to optimize possibilities for re-use.

The data owner will assign a version number from V0.1 to V0.9 to each dataset when the data is still being used (storage phase). The final deposited data will be given the version number V1.0 (archive phase). Any changes made to the dataset after deposition are logged in the repository's version control system.

Naming conventions

PLATO datasets will be named with project title (PLATO), creator/owner of the dataset (ESR1-ESR15), number of the dataset of this researcher, and a description of the type of data; thus according to the following formula:

→ PLATO-ESRX-00-datatype [i.e. PLATO-ESR3-01-interviews]

2.2. Making data openly accessible (A)

Data repository

After the end of the project (31/12/2020), and as soon as project results have been published in the form of peer-reviewed publications, all data generated by PLATO, including metadata, documentation and code, will be deposited in a certified national data repository which provides a support structure for data preservation and sharing as well as safe and sustainable data storage facilities. As a general rule, project partners will use the Dutch DANS data archive,[2] while partners may also choose to store in other repositories,[3] such as the

---

[1] http://www.ddialliance.org/

[2] DANS, the Netherlands Institute for Permanent Access to Digital Research Resources (https://dans.knaw.nl/en) is an institute of the Netherlands Academy of Sciences (KNAW) and the Dutch research funding organisation NWO.

[3] In the case of anonymised interview transcripts in other languages than English or that of the repository's country, a national archive may not accept the files as correct anonymisation is manually checked.

Norwegian Centre for Research Data (NSD).[4] Upon deposition, the data supplier and repository will sign an archiving agreement, which regulates the terms and conditions for archiving and making research data accessible.

Software

The individual datasets will specify if any particular software tools are needed to access the data. In the case of text files (notes, transcripts), the file format will be PDF and Adobe Reader will be necessary to access the data. In the case of other files; statistical data, audio or video, this will be further specified according to the file formats used.)

Open data vs. data with restricted access and/or use

The survey data generated by PLATO will be made openly available by default.

The data generated through semi-structured and focus group interviews will be treated with confidentiality and prudence to ensure no violation of regulations, legislation or ethical frameworks. Informed consent forms will be used for all interview participants.

It is expected that a majority of the semi-structured interviews will not be recorded due to sensitivity issues (i.e. policy areas under high media attention, personal values/attitudes) and the risk that interviewees would hesitate to speak frankly and openly about the questions at hand if recorded and answers later being made openly accessible. In these cases the researcher will take interview notes and write up summaries. Where the issues at hand are less sensitive, recording may be useful and full interview transcripts can be made available after anonymisation. Interview notes or transcripts will be made available for third parties to the extent possible, however taking into account the need to ensure full anonymity so that data cannot be traced back to individual persons, as well as a cost-benefit analysis of such anonymisation in each case.

As regards focus groups, personal and potentially sensitive data will not be collected. Each focus group will be moderated to ensure that participants can express their views without pressure or abuse. The discussions will be on public policies and procedures, and the purpose of the focus groups will be to collect different types of arguments. The different arguments will not be traceable back to individual persons, and the discussions will not be made publicly available, only a summary of the findings. All data will be reported at an aggregate level and no information can thus be traced back to individual respondents.

Access restrictions

If access restrictions apply to the data, data will be made available to others in accordance with conditions set out in the individual agreements between the data supplier and repository, and access will be given i.e. on the basis of a signed loan agreement.

Further conditions for access to the data will be specified in the next version of this DMP.

---

[4] NSD, the Norwegian Centre for Research Data is one of the world's largest research data archives: http://www.nsd.uib.no/nsd/english/index.html Their NORDi project (Norwegian Open Research Data Infrastructure) will develop tools to make it easier to locate, use and share research data.

Website

The PLATO website will make reference to the data (including metadata) generated which is available for further exploitation and link to the data repository where it is stored. In cases where full open access to the data is not possible due to confidentiality or ethical concerns as outlined above, overviews/protocols of data will be made available.

2.3. Making data interoperable (I)

Standards and formats

The data to be deposited and made available for other researchers will be archived in standard formats and in compliance with available open software applications to allow for data exchange and reuse. The file formats for the different kinds of data are listed in Table 1.

By using DCC as metadata standard (see 2.1 above) the interoperability with the European Open Data space is guaranteed.

2.4. Increase data re-use (R)

Licencing

Data will be made available under a CC licence, which manages copyright and the terms and conditions for access and reuse. A CC licence requires attribution of the licensor (data supplier) when the use involves public sharing.

After the end of the project (31/12/2020), and as soon as project results have been published in the form of peer-reviewed publications and/or PhD theses, PLATO data will be deposited and made available for third parties. The moment, at which the data will be made available, any conditions for re-use by third parties and embargoes after the end of the project will be specified in a subsequent version of this DMP.

3. Allocation of resources

3.1. Costs

Costs related to data management are expected to be limited and will be covered by the individual beneficiary's institutional unit costs. Each ESR will be responsible for setting up a budget for data gathering/collection with the help of local research and/or financial support staff. These could include travel and accommodation costs for the interviewer, and in the case of focus groups also for the interviewees, subscription costs for access to databases (i.e. press clippings, journals) and software tools.

Costs for the anonymisation of interview transcripts could be extensive; therefore a cost-benefit analysis will be made by the ESR where this is a relevant task. The costs will be discussed with the beneficiary (supervisor and research manager/financial officer) and the Coordinator and Project Manager at ARENA, who will make sure that the provisions of the PLATO Grant Agreement and this DMP are followed. The potential value of processing such data with the purpose of making them available for re-use by other researchers must be estimated as higher than the costs in terms of the work load, which will depend of the amount/size of the data, and sensitivity/confidentiality issues.

Costs related to the administrative coordination of data management and deposition in line with this DMP will be covered by project funds at network level administered by the coordinator.

Costs for the long-term preservation of data will not occur, as the archiving in the DANS data archive and other certified national repositories is free of charge.

3.2. Project responsibilities

Open Data Access is dealt with in WP8, where the lead beneficiary is the University of Twente. PLATO's Steering Committee 2 on Supervision, Management & Dissemination has particular responsibilities for elaborating and adopting the DMP, while PLATO's Project Manager has responsibilities regarding the establishment and development of the document, as well as information to all project participants and overseeing that the requirements are met and deadlines respected by the individual researchers. Each researcher (ESR) is responsible for creating an individual data management plan by means of the attached PLATO template, and for handling the research data in line with the provisions of the PLATO DMP.

4. Data security

PLATO's ESRs are employed at university or higher education institutions which all offer several options for secure storage of data. Throughout the project, in the data gathering and analysing phase, individual researchers will store data in their personal password-protected spaces on secure university servers, which provide appropriate back-up solutions.

At the end of the project, the data will be deposited for long-term storage in certified repositories, such as the DANS data archive.

5. Ethical aspects

Some of PLATO's researchers will conduct social science research that involves human beings as voluntary study participants in focus groups, interviews, and surveys. D7.3 outlines how the project will deal with research ethics in this field, including details on the procedures and criteria that will be used to identify and recruit research participants; ethics approvals; and templates of informed consent forms and participant information sheets.

Some researchers will generate or collect data from Non-EU countries (NEC), i.e. Macedonia. D7.2 confirms that the ethical standards and guidelines of H2020 will be rigorously applied also for research carried out in non-EU countries.

For the researchers processing personal data collected through focus group interviews, free and fully informed consent of the persons concerned will be ensured, as well as the necessary notifications/authorisations from the participants. The data will not include any sensitive personal data such as health, sexual lifestyle, ethnicity, religious or philosophical conviction, but political opinions may be part of the focus group interviews. D7.1 further explains how these issues will be dealt with in PLATO.

6. Template for individual datasets

Project identifier*	H2020-MSCA-ITN-2016-PLATO-722581
Project name*	The Post-Crisis Legitimacy of the European Union (PLATO)
Project abstract*	PLATO investigates the legitimacy of the EU's responses to the financial crisis. It uses the example of the financial crisis to build and test theory of what would amount to a legitimacy crisis in the case of a multi-state, non-state political system such as the EU. PLATO generates new understandings of the EU's legitimacy crisis through multiple, connected case studies undertaken by 15 Early Stage Researchers (ESRs). PLATO project flyer (pdf)
Reference*	**PLATO-ESR*X*-00-datatype**
Title*	*Name/description of dataset (i.e. Focus group interviews on [theme])*
Name of researcher*	*Full name of the creator*
Institution*	*Name of PLATO beneficiary*
ESR no.	*ESRX*
PhD project title*	*Title/topic of PhD project*
Abstract*	*Briefly describe the aim of the study and data gathered/collected*
Data type*	*Specify the kind of data generated (i.e. semi-structured interviews, focus group interview, survey data, see Table 1 of PLATO DMP)*
Data format*	*i.e. text, database, audio, video and file format*
Keywords*	*Describe the data and related analysis/research thematically*
Language*	*If different language at different levels/phases, include this information (i.e. one language used for recording/notes during interviews, English in processed notes/final archiving stage)*
Units of analysis*	
Geographical coverage*	*Countries/regions covered*
Time period covered*	*Start and end date of the data (DD/MM/YEAR)*
Description of data	Select either A) *New data generated* or B) *Re-use of existing data* below
A) New data generated	
Sample*	*Explain how the sample is made; population, criteria for inclusion/exclusion*
Collection period*	*Start and end date of data collection (DD/MM/YEAR)*
Other	*Any other relevant info to be added?*
Copyright*	*Insert own name (copyright belongs to the creator?)*
B) Re-use of existing data/documents	
Sample	
Data source	*What is the origin of the data?*
Copyright	*Specify whether you have the rights to collect and provide access to data*
	To be filled in for both A) and B)
Methods*	*Describe the methods used for data collection/sampling*
Person-identifiable data*	*Yes or no. Where relevant, provide details on anonymisation and required permissions*
Version*	*V0.1-0.9 = Versions of the dataset in the research phase* *V1.0 = Version for final archiving*

Size*	File formats and size *Example: 15 interviews, which last on average 60 minutes each. Text files with notes of 1-3 pages for each interview. Expected size of final dataset: 15 PDF files * 100 kB: in total 1.5 MB*
Access rights*	*Describe if access will be given to the archived data (NB. Open access is the default option).* *Justifications must be given for any access restrictions. What kind of information will then be openly archived (overviews, protocols etc.)?*
Publications*	*Reference to the scientific publication(s) where the data is used*
Costs	*If applicable: Estimate costs for making data available and reusable i.e. through the anonymisation of interview transcripts*
Storage (phase 1)	*When actively using the data during the project:* *Specify the institutional IT solutions used for storage. University networked fileservers are ideal for storage and is to be used for storage of all data (password protection, security, back-up solutions). Note that storage only on a local PC/laptop hard drive or external portable storage devices (i.e. USB drive) is not allowed. Note also that remote/cloud storage is not appropriate for sensitive data and their service level agreements must be studied before use.*
Archiving (phase 2)	*After the end of the project (31/12/2020), and when the peer-reviewed scientific article(s) using the data have been published, the data will be deposited with the Dutch DANS data archive.* *If you prefer other certified national archives, give details and explain why (i.e. for language reasons in the case of anonymised interview transcripts, where a national archive may not accept files in other languages)*
Other	*Any other information or comments*

* Metadata