1 DOCUMENT INFORMATION

Project	
Project acronym:	INFORM
Project full title:	CLOSING THE GAP BETWEEN FORMAL AND INFORMAL INSTITUTIONS IN THE BALKANS
Grant agreement no.:	693537
Funding scheme:	H2020-INT-SOCIETY-2015
Project start date:	01/04/2016
Project duration:	36 months
Call topic:	INT-10-2015
Project web-site:	https://formal-informal.eu/en/
Document	
Deliverable number:	D1.1
Deliverable title:	Data Management Plan
Due date of deliverable:	30 Sep 2016
Actual submission date:	31 Dec 2017
Resubmission date:	27 Jun 2018
Editors:	Dr Klavs Sedlenieks, Dr Eric Gordy
Authors:	Dr Eric Gordy and Dr Predrag Cveticanin
Reviewers:	N/A
Participating beneficiaries:	UCL
Work Package no.:	WP1
Work Package title:	Coordination and Evaluation
Work Package leader:	UCL
Work Package participants:	UCL,IEF,UM,CISAR,CESK,IDSCS,QKHA,SRK,RSU
Estimated person-months for deliverable:	1.6M
Dissemination level:	Public
Nature:	Report
Version:	1
Draft/Final:	Final
No of pages (including cover):	13
Keywords:	Data Management Plan

1. DATA SUMMARY

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

The overall objective of this project is to provide a comprehensive understanding of the current situation in the Western Balkan countries through meticulous, bottom-up study of the interaction between the formal institutions brought about by EU integration processes and the informal institutions existing in the Western Balkan countries, thus providing policy makers with the tools for closing the gap between the formal and informal institutions in these countries.

Our specific project objectives are:

1. Advancement of empirical knowledge regarding the transposition, implementation and enforcement of EU rules in the Western Balkan states (on the basis of analyses of legal systems, case studies, surveys, interviews and secondary data analysis);

2. Advancement of empirical knowledge on informal institutions in the spheres of politics, economics and everyday life in the states of the Western Balkans and the relationship between formal and informal institutions in these states (on the basis of case studies, surveys, interviews, ethnographic research with participant observation and secondary data analysis);

3. Development of theory regarding the institutionalization of EU laws and regulations under conditions defined by ineffective formal institutions (and the powerful presence of substitutive and competing informal institutions);

4.   Creation of practical political measures directed toward the institutionalization of EU laws and regulations based on empirical evidence regarding the relations between formal institutions generated in the process of Europeanisation of Western Balkan states and inherited and newly developed informal institutions in the spheres of politics, economics and everyday life;

5.   Dissemination of project results to members of the academic community (at the international and regional level), national offices for European integration in the states of the Western Balkans, DG NEAR offices and to the general public.

In order to achieve these goals we have developed a mixed method research design using the following quantitative and qualitative methods for data collection:

(1)   the F2F survey employing multi-stage, national proportional, probabilistic samples of 6,200 respondents:

(2)   36 months of ethnographic work (6 months in each of the WB countries);

(3)   220 semi-structured interviews with survey respondents (chosen on the maps of social space of these societies by SPAD 8.0 software);

(4)   30 interviews with policy makers (6 from EU and 4 from each of the 6 WB countries);

(5)   A number of case studies using both quantitative and qualitative methodologies;

(6)   Content and discourse analysis of legal documents and media reports (related to the informal practices and interplay between formal and informal institutions in SEE societies)

(7)   Secondary data analysis on relationships between formal and informal institutions in different WB countries, as well as issues related to corruption and inter-ethnic relations (in the spheres of politics, economy and everyday life).

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

The original data produced by INFORM include the following resources: Surveys, Interviews, and Ethnographic field reports

Surveys:

The survey instruments are composed of the questionnaire to be administered, available in English and in the languages of the research region (Albanian, Bosnian, Croatian, Macedonian, Montenegrin, Serbian) and survey results produced through SPSS and SPAD software and readable/manipulable using standard statistical software packages.

Interviews:

The interview protocols will be available as word processing files.

Ethnographic field reports:

In order to comply with professional ethics codes, the field report data will not be shared otherwise than as a part of analytical reports.

Will you re-use any existing data and how?

Data from previous surveys (Eurobarometer, ANTICORRP, Transparency International, countrywide and regional opinion surveys) will in some instances be reanalysed and used for comparison and confirmation of the survey findings produced by INFORM. All of this data is published and available through open access.

What is the origin of the data?

Survey results will be generated by the agency chosen by INFORM through open tender, using the survey instrument that has been generated by INFORM researchers. Interviews and ethnographic reports are produced by INFORM field researchers.

What is the expected size of the data?

Sample sizes are described in the section above. As for the size of data files, no file will be larger than 10 MB, and it is expected that the size of all data files collectively will be smaller than 2 GB.

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

The principal users of INFORM project data will be researchers working on related topics in Europe and all around the world.

2. FAIR DATA

Making data findable, including provisions for metadata

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)?

What naming conventions do you follow?

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

Data will be encoded using the conventions for Sociology specified in the DDI Codebook. Data will initially be made available through the INFORM project website (http://www.formal-informal.eu/en/). After the end of the project data will be held and made available through the UK Data Archive (UKDA).

Do you provide clear version numbers?

The Management Board has decided that only final versions of research data will be made publicly available, and that raw data and working versions will be circulated only internally.

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.

Categories and keywords are the standard referents employing the benchmarks developed through Data Documentation Initiative (DDI, version 3.2). The principal referent will be the metadata standards for Sociology, although as the project is multidisciplinary additional keywords may be taken from the standards for other disciplines (Anthropology, Economics, Political Science). All of these are available through the DDI Codebook.

Making data openly accessible

Which data produced and/or used in the project 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.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

All processed quantitative data as well as interview summaries will be made openly available. Unshared data will remain at the discretion of the party that has obtained it.

How will the data be made accessible (e.g. by deposition in a repository)?

In the first instance data will be made available through the INFORM website. After the project is completed data will be deposited in the UK Data Archive (UKDA). The UKDA will assign a PID.

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

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Text files will be accessible using any of the widely used software packages available for reading and manipulation of text. They will be downloadable and can be analysed using any type of QCA software. Survey data will be downloadable and suitable for analysis using any of the standard statistical software packages currently in use.

No special software that is not commonly in use is required to access the data.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

After the project is completed data will be deposited in the UK Data Archive (UKDA).

Have you explored appropriate arrangements with the identified repository?

This will be done in the final period of the INFORM project.

If there are restrictions on use, how will access be provided?

There are no restrictions on the use of publicly shared INFORM project data.

Is there a need for a data access committee?

The INFORM Management Board carries responsibility for issues related to data access. After the end of the project period responsibility for issues related to data access rests with the Project Coordinator.

Are there well described conditions for access (i.e. a machine readable license)?

Having in mind that the data will be made available under a Creative Commons licence we will be using the machine-readable codes which will be obtained using the CC license chooser tool.

How will the identity of the person accessing the data be ascertained?

It is not necessary to ascertain the identities of people accessing the publicly shared data.

Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

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

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Categories and keywords are the standard referents employing the benchmarks developed through Data Documentation Initiative (DDI, version 3.2). The principal referent will be the metadata standards for Sociology, although as the project is multidisciplinary additional keywords may be taken from the standards for other disciplines (Anthropology, Economics, Political Science). All of these are available through the DDI Codebook.

Text files will be accessible using any of the widely used software packages available for reading and manipulation of text. They will be

downloadable and can be analysed using any type of QCA software. Survey data will be downloadable and suitable for analysis using any of the standard statistical software packages currently in use.

No special software that is not commonly in use is required to access the data.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

We will not be using uncommon, specific ontologies or vocabularies, but standard ones defined in DDI codebook for Social Sciences. In fact, we have chosen to follow standards defined by Data Documentation Initiative (DDI) in order to make our data interoperable.

Increase data re-use (through clarifying licences)

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

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.

Data will be made available under a Creative Commons licence accessible to all users under the condition that INFORM is acknowledged as the source of the data in case of publication. The embargo period applies to the period before participants in the INFORM project release the first publications of project data.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

All publicly shared data will be made available for use by third parties. No restrictions are applied.

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

We expect data to be preserved in perpetuity in the UK Data Archive (UKDA).

Are data quality assurance processes described?

As part of descriptive metadata we will describe, for quantitative data, the procedures of sampling, data collection, testing of logical consistency of data, ways of coding of data (including refusals to answer, no responses and missing data); in the case of interviews ways of sampling, interview guidelines and interview procedures, while in the case of ethnographic field reports procedures for observation and documentation will be thoroughly described, enabling assessment of their accuracy and overall data quality.

Further to the FAIR principles, DMPs should also address:

3. 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).

Open Access costs are budgeted in the INFORM grant. If necessary it will be possible to apply for coverage of additional costs through the UCL Research Office.

Who will be responsible for data management in your project?

The INFORM Management Board carries overall responsibility for data management. Within the Management Board the individuals carrying principal responsibility are Eric Gordy (Project Coordinator), Predrag

Cvetičanin (Research Coordinator) and Klavs Sedlenieks (Outreach 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)?

As already stated the decisions on how what data will be kept and for how long will be made by the Management Board of the INFORM project, but resources for long term data preservation are not yet determined (these will be determined during the final year of the project). After the expiry of the project period, responsibility for data preservation resides with the Project Coordinator.

4. DATA SECURITY

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

All publicly shared data, in addition to being available through the INFORM website and appropriate repositories, are also held by the coordinating institution (UCL) and the institution coordinating the field research (CESK). In the event of data loss or corruption this form of triangulation allows for recovery of data.

The publicly shared project data includes no sensitive data calling for special measures related to security.

Data that may contain sensitive information (ethnographic field data and qualitative interview data) will be stored in an encrypted form in password-protected environments. Procedures will be devised to ascertain maximum separation of any identifiers and the data.

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

The publicly shared data will be deposited in the UK Data Archive (UKDA) for long term preservation and curation

5. 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 research proposal is being reviewed by UCL Research Ethics Committee. We have identified the main area in which particular sensitivity is required as the safety/privacy of those people subject to the research.

We will not be working with people who would be regarded as vulnerable by any standard definition, such as children. Most of the ethnographic portions of the study involve the collection of data based on the knowledge, attitudes and practices of adults that, while not necessarily personal, may be sensitive in the sense that people are persecuted for their political views in many parts of the world. This requires a critical commitment to the preservation of anonymity, not only in the final presentations of our data but also in the storage of this data prior to and after publication. The data made available to researchers will be anonymised data. Raw data will not be included in the project archive.

In practical terms, this requires ensuring that all fieldworkers are trained in methods of keeping and storing their field notes in formats (electronic or otherwise) that would not allow a third party to identify persons either from names or other distinguishing features such as job titles. We will work together with UCL Research Ethics Committee to ensure that we consistently abide by data protection concerns with respect to the safe storage of personal data.

The sample of research sites includes non-EU countries, some of which are lower-income countries. Although standard schemes of benefit-sharing do not apply, consciousness of inequalities in the relationship between international researchers and domestic publics forms an essential element of the ethos of ethnographic research. The experienced researchers on the project will be sensitive to power differentials inherent in this type of international research. Our research plan involves the engagement of domestic academics in the research as members of the project advisory board and as local advisors to the field researchers (as described above). We also intend to seek internal university funding for workshops in the host countries of the research at which findings will be presented and shared with the local academic and policy communities.

Because the project consortium consists of researchers from a range of countries, communication and coordination of ethical standards will be essential. Guidelines in compliance with EU standards and UCL procedures will be distributed to all researchers as part of the project coordination.

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

The survey and interview instruments include information to respondents regarding the anonymisation of data and data preservation.

6. OTHER ISSUES

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

In the final instance policies will be determined in compliance with the guidelines provided by the coordinating institution. The overseeing office is UCL Research Data Services (https://www.ucl.ac.uk/research-it-services/research-data-service), which assures compliance with procedures outlined by the UCL Research Data Policy (https://www.ucl.ac.uk/library/research-support/research-data/policies).