

From Leeds RoaDMaP Engineering training handbook available at:
http://library.leeds.ac.uk/info/377/roadmap/123/roadmap_events/2

[Annex B - Data Management Plan B]

ESRC-DFID Example Data Management Plan

http://www.esrc.ac.uk/images/Example-Data-Management-Plan_tcm8-20657.pdf

Existing data

The research objectives require qualitative data that are not available from other sources. Some data exist that can be used to situate and triangulate the findings of the proposed research (eg, surveys of poverty impacts; opinion polls), and which will supplement data collected as part of the proposed research. However, qualitative and attitudinal data are generally rare or of insufficiently high quality to address the research questions.

The research objectives also require quantitative analysis of public data. Some quantitative data are available, but they are insufficiently detailed. In their current form, they would not permit as full a comparison across the cases as is desirable.

Information on data

For these reasons, the research project involves primary data collection: 1) public data; 2) semi-structured interviews; and 3) focus group discussions with people identified through profiling techniques:

1. *Public data*

Where possible, we will use online and/or electronic archives. This will involve extracting and processing quantitative data, including participants, objectives and outcomes. Key search terms and their translation into the relevant languages, inclusion and exclusion criteria for items, variable codes and metadata will be refined and agreed in the inception phase of the project.

Preliminary searches indicate that a sufficiently detailed dataset can be generated. The junior researchers will log their progress, documenting potentially contentious categorising decisions, difficulties faced in categorising items, and qualitative insights which do not fit the spreadsheet format. Data will be inputted and stored in a widely available spreadsheet format (eg Excel or SSPS), to ensure accessibility to Southern researchers.

2. *Semi-structured interviews with individuals*

The team anticipates undertaking 25-40 semi-structured interviews in each country from a sample frame to be developed in Phase 2. Data will be collected and stored using digital audio recording (eg MP3) where interviewees permit. In case they do not, interviews will be undertaken in pairs to enable detailed note-taking. Interview notes will be typed up according to agreed formats and standards. Where interviews are taped and in English, the UK research assistant will assist with transcription.

3. *Focus group discussions matched to profiles*

The sample frame for the focus group participants will be derived from public data. Numbers of focus groups will depend on geographical and other variations in patterns; how quickly a robust pattern of findings emerges; and the scope for identifying and convening the appropriate groups. Focus groups will involve two researchers, and be conducted in the vernacular. Whether recorded or not, the event will be transcribed or documented using agreed formats and standards for handling the issue of multiple voices,

interruptions, labelling of participatory and visual activities, and so on.

All transcripts will be in Microsoft Word. All the researchers (except the UK research assistant) will be reasonably fluent in both English and the main language in which interviews and focus groups will be conducted, so that transcriptions will be translated into English only where the researcher is fluent in both languages and better able to transcribe in English, or to enable analysis of particular sections of the text. This will avoid unnecessary cost.

During the inception Phase 2, the metadata, procedures and file formats for note-taking, recording, transcribing, storing visual data from participatory techniques, and anonymising semi-structured interview and focus group discussion data will be developed and agreed. Focus group and interview transcripts will be coded in NVivo or a qualitative software suited to the different languages; the most appropriate software for a comparative multi-language study has not yet been identified.

Quality assurance

The PI will be responsible for overall quality assurance, with lead country researchers and the UK research assistant undertaking specific activities to ensure quality control. Detailed protocols for extracting data from secondary sources will be developed, piloted, refined and agreed in Phase 2.

Quality will be assured through routine monitoring by the lead country researcher, and periodic cross-checks against the protocols by the UK-based research assistant. While interview and focus group protocols are being developed in Phase 2, standards and systems for note-taking, recording (if possible), transcribing and storing visual data from participatory techniques such as drawings, photographs and video, use of metadata, systems for downloading and storing SMS data (a potential follow-up research tool) will also be defined.

Focus groups and interviews will always involve two researchers. Quality control for the qualitative data collection will be assured through refresher focus group discussion training during research design workshops and to junior researchers, where appropriate. Either the UK Institution or lead country researcher will check through each transcript for consistency with agreed standards. Where translations are undertaken, quality will be assured by one other researcher fluent in that language checking against the original recording or notes.

Backup and security

Our data will need to be backed up regularly; because of likely problems with viruses and hardware in developing countries, this will include regular email sharing with the UK research assistant, so that up-to-date versions are stored on the UK Institutions server. Qualitative data will be backed up and secured by the lead country researcher on a regular basis and metadata will include clear labelling of versions and dates. There are some potential sensitivities around some of the data being collected, so the project will establish a system for protecting data while it is being processed, including use of passwords and safe back-up hardware.

Ethical issues

A letter explaining the purpose, approach and dissemination strategy (including plans to share data) of the research, and an accompanying consent form (including to share data) will be prepared and translated into the relevant languages.

A clear verbal explanation will also be provided to each interviewee and focus group participant. Commitments to ensure confidentiality will be maintained by ensuring recordings are not shared; that transcripts are anonymised and details that can be used to identify participants are removed from transcripts or concealed in write-ups.

As the highly-focused nature of the research means that many participants may be easily identifiable despite efforts to ensure anonymity or confidentiality, where there is such a risk, participants will be shown sections of transcript and/or report text to ensure they are satisfied that no unnecessary risks are being taken with their interview data.

Interviews with elite policymakers will not guarantee confidentiality unless this is requested, as interviewees will be expected to speak in their official capacities or institutional roles. However, as is often the case, interviewees may be more comfortable if some sections of their interview are not

recorded or made public. In such circumstances, recording will be paused or sections of text will be expunged from shared transcripts, and an indication made that this is the case. □

Expected difficulties in data sharing

Not all of the transcripts will be translated into English (see above), which will limit the accessibility of the data.

Copyright/Intellectual Property Right

The institutional partners will jointly own the data generated. Online and archival sources will be cited and clearly acknowledged in the database and research outputs. Permission will be sought from secondary sources to share the findings of the research on public websites.

Responsibilities

The PI will direct the data management process overall, with the UK research assistant responsible for ensuring metadata production, day-to-day cross-checks, back-up and other quality control activities are maintained. The lead country researchers will be responsible for routine supervision of the dataset development.

Data extraction, processing and inputting for the dataset will be undertaken by the in-country junior researchers. The UK Institution, lead country and junior researchers will share responsibilities for collecting and transcribing focus group and interview data, with the UK research assistant supporting as necessary. The PI will be finally responsible for dealing with quality and sharing and archiving of data.

Preparation of data for sharing and archiving

The most appropriate means of sharing the data generated through the project will be online, through institutional websites. The project will have a dedicated space on the UK Institutional website to facilitate this, and all other involved institutions will also be encouraged to host the data on their websites.

[Annex B - Data Management Plan B]



Public relations, society and the public sphere
Project Stage: Application
RCUK Research Councils: Economic and Social Research Council
Lead Organisation: University of Leeds
Project dates: 1 January 2013 to 31 December 2015
Budget: £414,238.00

1 Existing data sources

1.1 An explanation of the existing data sources that will be used by the research project (with references).

DCC 2.2.2: What existing datasets could you use or build upon?

A number of data sets are named in the proposal, as follows:
National Census Data 2011; PRCA member survey 2011; CIPR member survey 2011.

2 Gaps between the currently available and required data

2.1 An analysis of the gaps identified between the currently available and required data for the research.

DCC 2.3.1: Why do you need to capture/create new data?

Existing quantitative data will provide enough material to establish the structures of the PR industry. However, there is no existing qualitative data about the ways in which the cultures and practices of PR across different contexts contribute to the shape of the field. For this reason, new qualitative data is required, targeting this information specifically.

DCC 2.4.1: What is the relationship between the new dataset(s) and existing data?

Existing data will facilitate a quantitative analysis of the field's objective structures (company sizes, specialist sectors, turnover, geographical spread, practitioner demographics). This will inform the analysis of the qualitative data and put into context the cultural and practical norms that characterise the field. The analysis of both aspects of the field will help clarify why certain meanings and values are attributed to different forms of practice across the field, and how these different attributions ultimately shape the industry as a whole.

3 Information on the data that will be produced by the research project

3.1 Data volume and data type, e.g. qualitative or quantitative data

DCC 2.1: Give a short overview description of the data being generated or re-used in this research

Existing quantitative data will be drawn from industry associations, the ESDS databases and government sources, in order to establish the objective structures of the field. These sources are:

- National occupational and demographic data including the 2011 census data
- Existing CIPR and PRCA membership data covering industry structures including sectors, specialisms, salary levels, educational data, patterns of tenure, practitioner demographics and

distribution across the field, consultancy size, turnover and locations (e.g. PRCA, 2011, CIPR 2010 and the annual PRCA benchmarking survey)

- CIPR and PRCA membership survey data
 - Quantitative analysis of PR award recipients (campaigns and practitioners) for the last decade (this database will be constructed from existing records)
 - ESRC databases (Miller, 2005)
 - The Great British Class Survey (BBC-led project, data will be publicly available in 2013)
- Qualitative data sources are new, raw observational data gathered from interviews, focus groups and existing documentary sources, and will comprise:
- Professional association and consultancy web sites and promotional material
 - Industry awards submissions and judgements
 - Interviews and focus groups with practitioners
 - Interviews with recruitment specialists in PR
 - Case analyses of specific issues as examples of 'public sphere' debates

3.2 Data quality, formats, standards documentation and metadata

DCC 2.3.3: Which file formats will you use, and why?

Microsoft Word 2007 for text based documents. MP3 or WAV for audio files. Quicktime Movie or Windows Media Video for video files. Quantitative data analysis will be stored in SAV file format (used by SPSS) from which data can be extracted using the open-source spssread Perl script. 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.

DCC 2.3.4: What criteria and/or procedures will you use for Quality Assurance/Management?

Existing quantitative data sources will be subject to the quality assurance processes of the organisation that originally gathered the data. However, we will review the data and ensure it is 'cleaned' for use in our own study. Where possible and permitted by participants, data will also be stored in the ESDS archive, as per the terms of ESRC funding.

DCC 2.5.1: Are the datasets which you will be capturing/creating self-explanatory, or understandable in isolation?

Yes

DCC 2.5.2: If you answered No to DCC 2.5.1, what contextual details are needed to make the data you capture or collect meaningful?

However, in order to be absolutely clear about the meaning of the data, each data set will have an overview document attached to it, covering the details required by ESDS that will facilitate their use by other researchers (e.g. sample structure, focus of the data collection, dates of collection, etc).

3.3 Methodologies for data collection

DCC 2.3.2: Describe the process by which you will capture/create new data

The following methods will be used to collect new data:

1. Collating existing web site and promotional material from professional associations and consultancies;
2. Collation of submissions to and judgements of industry awards;
3. Interviews with practitioners (60 in total), PR specialist recruiters (5); end-users of PR (15)
4. Focus groups with practitioners in different UK regions (6)
5. Case analyses (3 issues) comprising PR-driven documents created by stakeholder

organisations in relation to the issue; and media coverage of the different perspectives on the issue (including PR-driven and non-PR driven media coverage).

4 Quality assurance and back-up procedures

4.1 Planned quality assurance and back-up procedures (security/storage)

DCC 5.2.1: How will you back-up the data during the project's lifetime?

Electronic data will be stored on the University of Leeds SAN (Storage Area Network), which comprises enterprise level file servers in physically secure data centres with appropriate fire suppression equipment. Snapshots are taken every day at 10pm (and accessible for 1 month). A second level of snapshots is taken every month and are kept for 11 months. Snapshots are user recoverable from the desktop.

An incremental copy to backup tape is taken every night (and kept for 28 days) and a full copy is taken every month. Every quarter, the full dump tapes are moved to a long term storage facility where they are kept for 12 months.

Tapes are initially stored in on-campus fireproof safes and then moved to off-campus secure locations.

DCC 5.3.1: How will you manage access restrictions and data security during the project's lifetime?

Access to electronic data is controlled by Active Directory (AD) Group membership. The Faculty IT Manager will set up a dedicated folder for this research project and create read-only and read-write AD groups. The PI will decide which users require read-only and read-write access.

Off-campus access is via the Citrix portal.

External users who need access to the data will apply for a University username and then be assigned to the appropriate AD group.

DCC 5.3.3: Give details of any other security issues.

Any sensitive data (as defined by the Data Protection Act) that is stored on portable electronic devices will be protected by encryption software to FIPS 140-2 standard. Any sensitive data that needs to be transmitted electronically will first be encrypted to FIPS 140-2.

If any highly sensitive data needs to be stored, then research data folder on the SAN will be encrypted, so it can only be accessed by authorised members of the project with the appropriate encryption software installed on their desktop PCs. Highly sensitive data is not available from off-campus.

5 Management and archiving of collected data

5.1 Plans for management and archiving of collected data

DCC 6.1: What is the long-term strategy for maintaining, curating and archiving the data?

Data which is able to be made publicly available will be offered to the UK Data Archive at the end of the project.

6 Difficulties in data sharing

6.1 Expected difficulties in data sharing, along with causes and possible measures to overcome these difficulties. (You may wish to include explicit mention of consent, confidentiality, anonymisation and other ethical considerations.)

DCC 3.1.1: Are there ethical and privacy issues that may prohibit sharing some or all of the dataset(s)?

Yes

DCC 3.1.2: If you answered Yes to DCC 3.1.1, How will these be resolved?

Some consultancies or individuals may be reluctant to have their data shared. Such concerns will be resolved for the interviews and focus groups by ensuring all transcripts are anonymous and identifying detail (e.g. campaign names) are removed or given pseudonyms in the transcript.

The case analyses, industry awards data and analysis of promotional material will proceed on the basis of publicly available documents, so the issue of sharing should not arise. However, copyright may be an issue for some of these documents. To resolve this, and before data is stored by ESDS, copyright permission will be sought for web and other promotional material written by participating organisations.

7 Copyright and intellectual property

7.1 Copyright and intellectual property ownership of the data

DCC 3.2.1: Will the dataset(s) be covered by copyright or the Database Right? If so give details in DCC 3.2.2, below.

Yes

DCC 3.2.2: If you answered Yes to DCC 3.2.1, Who owns the copyright and other Intellectual Property?

See response to question 3.1.1

8 Responsibilities for data management and curation

8.1 Responsibilities for data management and curation within research teams at all participating institutions

DCC 7.1: Outline the staff/organisational roles and responsibilities for data management

The PI will have overall responsibility for implementing the data management plan. The Faculty IT Manager will be responsible for ensuring that electronic file permissions have been correctly assigned and for advising on other aspects of data storage and security. Staff involved in the project at participating institutions will be responsible for following data management procedures.