

## **Data management plan**

### **Existing data and the need for new data**

A range of quantitative data exists on Chinese engagement in African countries (accessible through sources such as Chinese Investment Tracker, AidData and OECD FDI flows statistics), and this will be used for background as indicated in the Case for Support. Much of this data is, however, highly problematic, and to interpret it well I will benefit from the support of IAG member Deborah Brautigam, who is globally renowned as a critical analyst of existing data on China-Africa relations, as well as from the expertise of mentor Graham Harrison on working with donor-based data more generally. Aside from this basic background data, there is very little existing data capable of providing answers to this study's research questions. A range of qualitative data generated through existing research on China-Africa relations will be thoroughly examined, and discussed in Workshop 1 at the start of the project; however, to investigate the specific issues of Chinese influences on urban land policy, job creation, housing and mobility of the poor, the generation of new data for this project is essential.

### **Data generation and ethical considerations**

This project will generate a range of types of data. For the purposes of clarity in data collection, analysis, management and storage, the following categorisation of data formats is used (with approximate numbers in brackets): domestic elite and policymaker interviews (20-30 in each city); international donor and investor interviews (10 in each city); migrant businessperson interviews (10 in each city); local community survey responses (100 for each city); local community focus groups (1 in each city); cumulative site analysis interviews (6 participants per city, interviewed on several occasions); and field notes from 'techniques of following'. This data collection is achievable in the time allocated, because some participants will be interviewed by myself and some by the PDRA and there will be research assistance for survey and focus group implementation. I also base this on experience, having interviewed around 200 people during the course 6 months of field research for my PhD.

Survey and community focus groups participants will be assured of strict confidentiality, and the casual research assistants facilitating focus groups will be required to sign confidentiality agreements. Interviews are expected to average around one hour in length. Participants will be provided with a briefing sheet explaining the purpose of the research and making it very clear that all interviewees will be anonymised unless they explicitly agree otherwise, after discussion of the research purposes. Based on experience of similar kinds of interviewing, some interviewees may actively want their name to be used, while others may suggest that they do not mind it being used. In such cases it may strengthen the analysis to be able to identify people who agree to this, at least in terms of their organisational affiliation and position. It is, however, very important to make sure that such interviewees understand what the research is being used for, that the data may be archived for sharing and that they have considered any implications of being named or made identifiable.

All interviewees will be asked to sign a consent form stating they agree to be interviewed, and whether they are willing to allow the interview to be recorded. My experience in some of the research countries, and the sensitive nature of some of the issues to be discussed (for example in terms of critical comments on donor or migrant business practices, or alleged illicit deals between government officials and investors), strongly suggests that not all interviewees will be happy to be recorded - even if they are happy to be identified. There also may be cultural and cross-cultural issues that affect willingness to be recorded. This poses particular challenges for data management (including transcription) and sharing. These are, however, problems I have encountered extensively before. They will be mitigated by detailed note-taking, careful typing up of notes, and the provision of clear metadata for the purposes of data storage and sharing that indicates where interviewees were not willing to be recorded (see below). During field research, notebook pages will be scanned or photographed regularly as a backup, with the electronic copies kept in a password-protected folder on my laptop (and transferred to the secure university server when I return to the UK). The notebook will be kept in a locked safe when not in use.

## Data management and storage

The University of Sheffield data server will provide the basis for an electronic filing system for all data created, with folders being accessible only to myself and the PDRA. These files will be backed up both on my own computer and an external hard drive with password security. To ensure effective management of a large and varied corpus of data across three research countries, a detailed filing system will be used. This will utilise folders organised according to the three countries and the data types discussed above, and the careful versioning of files with dates and version numbers to ensure that there is no confusion over files in the process of transferring data between researchers and computers. Transcribed interviews (including as many of the cumulative site analysis interviews as possible, given the above) will, along with policy documents analysed under cluster 3, form the basis for critical discourse analysis using *NVivo* software.

Research consent forms will also be scanned and saved in password-protected folders. Any paper-based data gathered that is not in the public domain, as well as signed paper copies consent forms, will be kept in a locked filing cabinet. Unanonymised interview transcripts printed for the purposes of coding and analysis will be shredded after use, leaving only secure electronic versions. The PDRA will be coached on the importance of careful practices of data management and storage, to ensure that all files are kept in password protected files on the laptop purchased for his/her use, that any notebooks used are backed up and stored carefully, and that the versioning system is strictly adhered to.

## Data sharing

Given the increasing interest in Chinese engagement in Africa, speculation around the nature of aid flows, investments and their impacts – as well as a likely increase in studies concerned with how this is affecting African cities specifically – there is likely to be considerable interest in this data for subsequent research. The data generated will, wherever possible, be archived via the UK Data service in a timely manner at the end of the research programme. Given the above issues regarding anonymity and recording, there may however be complications with regard to sharing some of the data, as discussed below.

Survey data from the small-scale surveys conducted in each city will be uploaded, and as no survey participants will be identified and the questions are generic this should not pose problems for anonymity. The qualitative data collected may, in some cases, pose some more complex issues for sharing. Where interviews have been transcribed with anonymity requested and there is no foreseeable risk of the participant being identified through the interview content, transcripts will be uploaded to this service and listed using participant codes. In some cases, however, it may be necessary to remove certain sections of the interview itself where this jeopardises participant anonymity. Any passages removed in this way will be clearly identified. There may also be some interviews that cannot be shared because the content will too easily identify a participant who has asked to remain anonymous. In these cases, detailed metadata provided alongside the data archived will indicate the number of interviews not included, providing reasons why in each case. Where interviews have not been recorded and transcribed (for reasons discussed above), the metadata will again provide an index of these and information on how to access the relevant notes if required. Summaries of non-transcribed interview data can also be provided for sharing. Metadata accompanying the archived data will explicitly refer to the difficulties in comparing transcribed and non-transcribed interview data and ensure full clarity regarding which interviews have been shared, which have not, and why.

Participants' own interview responses, when transcribed, will be made available to them for corroboration where requested, as will data generated through the cumulative site analysis for those who participated in this.

## Intellectual property and responsibilities

The copyright of the data will be held by myself as PI, and institutionally by the University of Sheffield. Responsibilities for implementing this data management plan rest with me, as well as the PDRA with regard to specific aspects of the research as indicated above.