Data Management Plan

Our data management plan (DMP) aims to ensure that the data generated through this project is created, stored and made accessible in a shareable format. This will enhance the quality and rigour of the research, and maximise its impact. In preparing this DMP we have worked with the Digital Curation Centre's guidance documents and consulted our institutions' data and intellectual property units. If the application is successful, we will use the DCC's online planning tool (https://dmponline.dcc.ac.uk/) to prepare a comprehensive DMP.

Existing Data and Gaps Identified: The proposed project concerns international distance education (IDE) students registered for undergraduate, postgraduate and doctoral degrees in South Africa. There has been one comprehensive study on international students here but none on distance learners. Moreover, there has been no in-depth analysis on access to distance HE of students in the 3 major 'sending' countries – Lesotho, Nigeria and Zimbabwe- the focus of our study. A search of UK Data archive and South African electronic training and development database suggests that there is no equivalent dataset. Secondary data on demographic variables will be collated and analysed from existing institutional databases for the first time. New survey data and qualitative data will be created to understand the role of IDE in South African HE and to explore issues of equity and access. Existing public data sources such as data collected by the African Network for International Education (ANIE), SARUA and IEASA will also be consulted. The project requires access to data currently stored by the UNISA Directorate: Information and Analysis (DIA), which is retrievable in shareable format through the DIA's porthole to UNISA team members.

Information on New Data

Data Collection

The project will produce a mixture of qualitative, quantitative and learning analytics data.

<u>(1) Qualitative data:</u> Conducted online using video conferencing software. Face-to-face interviews will be undertaken with elite interviewees. Interviews will be conducted by project Research Associates (RAs) unless circumstances require participation by the Principal Investigators (PIs). They will be undertaken in English, the language of education at UNISA. Interviews will be recorded online and their contents transcribed. Interviews will be undertaken with the following cohorts:

- Up to 20 interviews with national and international policy makers, 5 from each of South Africa, Lesotho, Zimbabwe and Nigeria;
- Up to 15 interviews with academics from 3 STEM programmes;
- Up to 5 interviews with administrators involved in internationalisation policies;
- 180 interviews (20 mins each) with students in the final two years of study on undergraduate and postgraduate taught courses to explore issues of access, equity and quality, barriers to local learning, reasons for choosing South African HE, issues and challenges during the study, and future labour market and migration plans;
- Up to 5 representatives of new entrants to international distance education.

<u>(2) Quantitative data:</u> Quantitative data will be collected to compliment the qualitative data:

- A total of 1,100 questionnaires will be conducted amongst students in Lesotho, Zimbabwe and Nigeria, weighted according to overall national participation in IDE. 100 questionnaires will be undertaken with South African students for comparison purposes. They will provide data on access, equity and demographics, aspects of entry, retention, progression and the challenge of mobility. These surveys will provide descriptive quantitative analyses;
- Course-wide student outcomes will be collected and analysed for students in the final two years of undergraduate and taught postgraduate study. Changes to pedagogy will be undertaken, and matching data will be collected from the 2018 cohort.

Privacy

Consent forms with tiered data sharing protocol will be developed to allow research participants to accurately set out how they will and will not permit their data to be shared.

Data Management Plan

Data collected from individuals without a pre-existing public profile will be routinely anonymised to remove references to identity and place. Anonymisation of data collected from individuals with a public profile will be decided on a case by case basis. Interviewees will have the opportunity to speak anonymously if this facilitates data collection. Advice on how to deal with learning analytics data in a sensitive ethical manner will be provided by co-I Prinsloo who has published on this topic.

Data Analysis
Qualitative interview transcripts and field notes will be analysed using NVivo 10. Quantitative data from questionnaires will be analysed using SPSS. The central files for each dataset will be stored simultaneously on the OU's secure Hulse server and on the UNISA Unidrive server. Early in the project, a preliminary coding system for the qualitative data will be collectively devised which will be modified to meet the needs of the research as it develops.

Data Management
During the project: Qualitative interview data will be transcribed using trusted transcribers who have been employed in other projects at the OU and UNISA. Any additional field notes collected will be transcribed by their authors. All data stored electronically will, where possible, use open source and non-proprietary file formats to ensure ease of data sharing between platforms and longevity of data. Text data from interview transcripts and ethnographic field notes will be stored in RTF format. Qualitative data will be stored in CSV format. Good data management practice will be employed from the outset, with folder structures and file naming conventions devised according to good practice guidance from JISC Digital Media to ensure accessibility during the research project.

Quality assurance: All the interviews will be conducted in English and so cross-checking the transcript against the recordings is not difficult. By keeping the survey in-house we will be assured of good quality data. The Academic and Professional Advisory Group members will be asked to review the survey strategy and ensure that quality issues are being addressed.

Backup and security: All data collected will be stored on the OU's secure Hulse server and the UNISA UniDrive server and will be deposited in a secure area dedicated to this project. Automatic backups of this server are made every 24 hours. Data security will be maintained by ensuring research material is accessible only to the PIs and RAs. Off campus access will be possible using a secure VPN connection. All desktop and laptop computers where research material is accessed will be protected by passwords. If file transfers are required between research staff, either encrypted memory sticks or the secure online file transfer system ZendTo will be used. Thus, these protocol meet the U's Information Security Policy and the Data Protection Act.

Ethical issues: The research will be informed by the ESRC's principles, particularly integrity, honesty, confidentiality, voluntary participation, impartiality and the avoidance of risk. There is one area of ethical concern in this project. As the project involves interviews and questionnaires with students talking about their educational experiences it is important that they don't feel compelled to participate or feel that they have to post a positive picture about their institution. At the beginning of the questionnaire/interview assurances will be given to participants that they can withdraw from the study at any time and that any data collected thus far will then be destroyed. Moreover, great care will be taken to ensure that the association between participating students and the modules cannot be made in any published material or verbal presentations and this will be iterated during the interviews. Learning analytics data will also be collated and used so that no individual can be identified. This is important because the information they provide may be deemed 'sensitive' under the ESRC's Framework for Research Ethics (ESRC 2012: 9). We will assure participants about the anonymity of all data generated, during the analysis and in presentation of the data. Key persons who hold strategic information (e.g. ministry of education personnel) will also be assured of confidentiality if they so require, although it is recognised that this may not be necessary as they will be talking in their official capacity. Key informants will also have the opportunity to view and approve interview transcripts.

Data Management Plan

Storage: It will be made clear during all phases of the project that the data collected will be archived and made useable in the future by researchers not directly linked to the team that collected it. Consent will be obtained accordingly. All data deposited to the UK Data Archive will be anonymised. In South Africa all data will be classified and archived on the UniDrive at UNISA and this will be managed by Gunter. This server allows for multiple access from different locations at different levels. Thus data stored on the UniDrive can be accessed by all investigators remotely but only edited and manipulated by individuals with authorisation.

Copyright/Intellectual Property Rights: Much of the data generation involves primary data, but whenever access is required to databases and surveys, permissions will be negotiated with the owners for noncommercial, research, teaching purposes and dissemination activities. A one year embargo will be placed on data at the end of the project to allow analysis and presentation of original data. The project team will have access to specialist IP and legal support at The Open University and UNISA, which will be able to guide and support the team regarding IP matters for the entire project duration.

Responsibilities: The whole team will be responsible for data collection and to ensure good practice they will all be versed in the Data Management Plan and its technical and legal issues. We will do this through the first planning meeting prior to data capture. Data management at a strategic level will be the responsibility of Gunter. Data preparation for sharing and archiving will be the responsibility of Co-Is with assistance from RAs. This will occur in months 6-27 when most new data is generated. In the first months of the project, UK based PIs and RAs will attend a training seminar in research data management presented by the OU's Research Data Management Librarian. RAs will be responsible for day-to-day research data management including creation of documents using appropriate file formats, the naming of files and folders, and the generation of metadata where this is required. This will be overseen by the 2 PIs, who hold ultimate responsibility for data management.

Preparation of data for sharing and archiving:

1. The institutional data drawn from the DIA's server will be placed on the UniDrive with restricted access to all. All staff have editing rights to the questionnaire data and Prinsloo, Rienties and one RA to the learning analytics data.
2. Quantitative questionnaire surveys will be entered into Microsoft Excel and an application made to the UK Data Service (http://ukdataservice.ac.uk/).
3. Qualitative data from semi-structured interviews will be transcribed, coded using NVivo and once the data has been analysed and findings published, anonymised transcripts will be deposited also with the UK Data Service. Raghuram was Co-I for an ESRC project whose data is at the British Library's oral history archive and for teaching at the University of Exeter. She therefore understands the system and protocalls.
4. All archived data will be linked to peer-reviewed publications arising from this research, which will be registered and archived on the Open University's open access institutional repository – Open Research Online (ORO) at http://oro.open.ac.uk. ORO is one of the largest repositories in the UK. The site receives an average of 40,000 visitors per month from over 200 different countries and territories and has received over 2.5 million visitors since 2006. It enables access to research outputs via common search engines including Google, by using the OAI (Open Archives Initiative) Protocol for Metadata Harvesting.

Potential Users: As the project is being conceived, executed and disseminated along with key educational organizations working in this field – the APAG members – they will highlight the availability of this data to potential users. The publication strategy for academic and non-academic users and the project website will also enable potential users to familiarise themselves with the data. Potential users of the data set include future researchers although the findings will be of interest to academic INGOs and to government bodies.

Expected Difficulties in Data Sharing: Following the comprehensive plan set out above will facilitate data sharing. The project will make use of no commercially sensitive data and all personal data will be anonymised. As such, no difficulties are expected in data sharing.

Data Management Plan