Plan Overview

A Data Management Plan created using DMPonline

Title: Beyond Access: Entendiendo el acceso y uso digital en el sur global

Creator: Roxana Barrantes

Principal Investigator: Roxana Barrantes, Roxana Barrantes

Data Manager: Roxana Barrantes

Affiliation: IDRC

Template: DCC Template

ORCID iD: 0000-0002-7589-2272

Project abstract:

Several studies have shown the important role played by Information and Communication Technologies (ICT) in promoting development, most of them analyze how access to ICT allows rapid communication and transmission of information (access), finding positive contributions in a wide variety of contexts. This has led to optimism about the impact of these technologies, as well as increasing attention and policy efforts related to ICT access. Nevertheless, public policies focused in the access dimension may be insufficient: making technologies available to households and individuals also brings the need of promoting skills to take full advantage of the benefits ICT could provide. On the other hand, digital economy has experienced rapid growth in recent years, with the along with the increase in ICT use, becoming a central part of the everyday life of people, firms and governments. Through the great diversity of digital devices present at home, workplace and even public spaces, the use of ICT has transformed traditional ways of engaging in activities related to commerce, labor, transportation, education, health, social interactions, among others. However, this has not been always a social inclusive process: as more people become connected and able to enjoy the benefits of using ICT in their lives, the gap between those connected and not connected increase. This problem is not solved through connectivity-focused policy alone, the disparities evidenced in the intensity of use and in the unequal capabilities to reap ICT benefits should be considered. In this sense, Beyond Access project gathers individual information, not only about ICT access, but also about perceptions and different forms of ICT use, including Internet (e-learning, e-government, and gig economy jobs), mobile applications, social networks, among others. Beyond Access project surveyed people in six Latinoamerican countries: Argentina, Colombia, Ecuador, Guatemala, Paraguay and Peru. Information gathering took 2017 for Argentina, Colombia, Guatemala, Paraguay and Peru; and 2018 for Ecuador, and includes information about 9170 ICT users and non-users.

ID: 48884

Last modified: 24-02-2020

Grant number / URL: 108336 - 003

Beyond Access: Entendiendo el acceso y uso digital en el sur global

Data Collection

What data will you collect or create?

The project will generate a range of data types and formats. In particular, the project worked with qualitative and quantitative data. Thus, the data described in this plan will be shared, considering differences between both kind of data. Specific outputs and data sources from the project covered by this plan are: (1) Data Management Plans for each kind of data (qualitative and quantitative); (2) recorded interviews and transcripts (qualitative data); (3) survey responses from respondents of six Latin American countries part of the project (quantitative data); and (4) notes about the process of the project and main changes of the data. In addition, the project will generate at least three specific data formats that can be used through three software. Final versions will be shared, so that no further changes will be needed. The main forms of data that will be generated are: (1) spreadsheets (survey variables and indicators, database and records); (2) documents (focus groups transcripts as well as technical notes); (3) audio recordings (recordings of focus groups); (4) codes (specific format in Stata/SPSS software). Spreadsheets will be maintained as Excel or GoogleDoc formats and exported to CSV and DBF for data deposition and further analysis. Documents will be maintained as Word or GoogleDoc formats and exported to RTF for data deposition, or PDF. Audio files are maintained in a range of formats and will be deposited in an open format, specifically mp3. Codes will be maintained in Stata (dta) and/or SPSS (sav) formats, main software used by data analysts. Finally, as explained above, the data to be shared will be in its last version: data files will be fixed and not subject to substantial editing.

How will the data be collected or created?

It is not planned to use a formal standard or related tools to collect and create the data. Files of the project will be organized into general folders by type of data: qualitative and quantitative. Within the folders the files are named with dates and further relevant information (such as codebook, dictionaries and technical notes). It is important to mention that quantitative data will be organized in a single file that contains the information about respondents of six Latin American countries; while qualitative data will be organized in four different folders by country (Argentina, Ecuador, Guatemala and Paraguay), because of diverse contexts and questions asked in focus groups. Regarding versioning, we will use final versions of each dataset, spreadsheet, document, audio file, etc. Therefore, data files will be fixed and not subject to changes later. Finally, quality assurance will be provided by DMP and IDRC specialists (we will include the comments of reviewers of data repositories).

Documentation and Metadata

What documentation and metadata will accompany the data?

Data will be kept in standard and open formats (spreadsheets, CSV, DBF, audio records, transcripts in documents, Stata and SPSS) so it should remain readable for any interested party. However, some metadata (questionnaire, technical notes, variables dictionary, etc.) will be necessary. On the one hand, for quantitative data, a dictionary will be generated, including variables definitions, questions specifications, as well as alternatives for answers and associated labels. In addition, a document with technical notes will be generated in order to specify statistics methods for sample design. On the other hand, qualitative metadata includes all the details of the focus groups that were conducted. All documentation and metadata will be generated in a Document (Word, GoogleDoc or PDF) and published into a repository with the files described above.

We plan to use the Data Documentation Initiative (DDI) standard, in particular the DDI Codebook standard (since our dataset has only crossectional information). In addition, we will respect FAIR (Findable, Accessible, Interoperable, and Reusable) principles, such as creating accessible microdata and metadata related, and getting a Digital Object Identifier (DOI) for the project data.

Ethics and Legal Compliance

How will you manage any ethical issues?

When conducting both quantitative and qualitative data collection, consent was required from respondents; they were informed about the purpose of the research and about the use we would make of their information. In addition, datasets and all relevant

information will be anonymized. Moreover, data output to be shared does not include personal information as names, addresses or phone numbers. For both, qualitative and quantitative data, we will use identifiers to keep the identity of the respondents anonymous. Finally, the information gathering process (surveys, interviews and focus groups) was developed considering ethics recommendations and includes formal consents with participants. These consents indicate that personal information will be analyzed by our research team, and publicly shared if and only if data does not include personal information. In addition, we would follow recommendations of the data repository institutions where we would publish the data.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

The ownership of the data generated in the project is entirely ours. As part of the grant by IDRC, we have property rights and IPR to publish and share the data generated in the project. Moreover, IDRC encourages making the data public. Regarding licence for data reusing, we will follow recommendations and requirements for data repository. We are planning that once the project data is published in a data repository, there will be no restrictions to reuse data for any interested individual or institution. Finally, we will not seek any patent to publish/share our data. Deadline times indicated by IDRC for the publication process will be respected.

Storage and Backup

How will the data be stored and backed up during the research?

We anticipate less than one gigabyte of data and documents to be generated by the project. As far as possible data will be deposited in long term availability archives such as the IDRC Digital Library, the Open Science Framework (OSF), and the Internet Archive. Deposition will occur at the end of the project or when a relevant formal project output is published. Data and documents will be stored internally by our research team through a Google Drive folder. The research team, relevant members of the IDRC team, will be granted access to the Google Drive folders (no additional IT resources will be required).

How will you manage access and security?

The overall risk posed by these data are relatively small. The main risk is related to publishing personal information of the respondents and participants of the project. We will be very careful when generating the final output of the data that will be published in the data repository, not to include critical personal information in both, qualitative and quantitative data.
In addition to the research team, Google Drive folders (described above) will be shared only with members of the IDRC team. Therefore, only few people from two different institutions (team project and IDRC institutions) will be able to access and guarantee responsible use of the data. Furthermore, access to sensitive data such primarily audio (in Spanish only) and the related transcripts (in English) is maintained under access control through the Google Drive folders, which is only available for the research team and relevant members of the IDRC team.
Finally, we have already created a database (qualitative and quantitative) for the project. This was transferred into personal computers and software to analyze and generate data to be shared.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

We will only retain internally personal information of survey respondents (quantitative data) and participants of focus groups (qualitative data). In addition, we will have an internal meeting (workshops) with the members of the research team to decide what other relevant data to keep. On the other hand, data generated by the project could be used by researchers of different backgrounds (in particular, social sciences) and could be used through mixed methodologies (quantitative and qualitative analysis). Finally, project data will be opened to everyone (regarding individuals and institutions) who is interested once the data has been published in data repositories. Project data will be preserved in the repositories and in a Google Drive folder internally.

What is the long-term preservation plan for the dataset?

Data of the project will be held in the IDRC Digital Library, and Harvard Dataverse, Zenodo or UK Data Service. These data repositories will charge a fee depending on the size of the data (and related documents) that will be published. Finally, there is not any cost related to time and effort to prepare the data: the research team has already created a database (qualitative and

quantitative) for the project.

Data Sharing

How will you share the data?

We will use the project web page and social networks, as well IDRC social networks to post information about the publication and updates of project data. We will share project data with any individual or institutions who want to use and analyze the data, but enforcing intellectual property rights through creation of a standardized formal text for citation purposes, and demanding a correct citation and use of the data and authors' referencing. Therefore, we will use a Creative Commons licence for non-commercial activities.
By May 2020, we will publish the data via a data repository (including a DOI identifier), but we request information to data repositories managers about downloads and any other information requirement (sampling methods, particular questions of qualitative data collection, etc.).

Are any restrictions on data sharing required?

The project database will be adapted to the requirements of the repository. In case the restrictions do not have solutions, we will handle other repositories alternatives, as we mentioned above. We do not need exclusivity for using the data, as we agreed with IDRC (we only need a data sharing agreement with the data repository).

Responsibilities and Resources

Who will be responsible for data management?

The DMP will be initially developed by Diego Aguilar (Research Assistant), and further developed by Roxana Barrantes (Project leader) and Aileen Agüero (Project coordinator and researcher), main responsible for implementing the present DMP. In addition, this DMP will be reviewed by IDRC Open Data Research Initiative team. Data curation and preparation to publication is part of the task of Diego Aguilar. Metadata generations will be carried out by Aileen Agüero. Finally, the Data Article will be developed by Roxana Barrantes. We plan to take advantage of the meetings and workshops organized by IDRC, as part of the Open Data Research Initiative program. Finally, Contract agreements for DMP activities are only under responsibility of our research team.

What resources will you require to deliver your plan?

We do not require any additional resource or specialist expertise nor additional or exceptional hardware or software requirements. Data repositories charge a fee to publish data and associated documents (e.g. metadata).