

Data Management Plan: HarassMap

Reem Wael ‡

‡ Harassmap, Cairo, Egypt

Corresponding author: Reem Wael (reem.wael@harassmap.org)

Reviewable v1

Received: 18 Jul 2017 | Published: 18 Jul 2017

Citation: Wael R (2017) Data Management Plan: HarassMap. Research Ideas and Outcomes 3: e15133.
<https://doi.org/10.3897/rio.3.e15133>

Abstract

HarassMap is an Egyptian organisation that works to create an environment where sexual harassment is not tolerated, and where individuals and institutions take action against it. For the purpose of this project, the project team cleaned up, organised, and made openly available for the public to access and use through a web portal, three main types of data:

- Crowdsourced reports of sexual harassment incidents (reports on HarassMap's online reporting and mapping system) - CSV and XLS
- Field data from HarassMap's research on sexual harassment using traditional qualitative and quantitative research methods - DOCX, PDF, SAV, MP3
- Social media conversations (comment threads and messages related to sexual harassment on HarassMap's Facebook page) - XLS

The social media data was collected retrospectively from our Facebook page during the project period and covers the period 2010-2016. The crowdsourced data and the research data was cleaned and organised to make sure it is usable for the public but still kept in its raw format. During the collection and organisation period, we also made sure to clear out all personal identifiers from the data to ensure anonymity and confidentiality, and prepared descriptions of each dataset that will help the public understand how the data was collected and how it can and cannot be used.

The data is stored online on a web portal that we built together with a web developer during the project period. On the web portal, the data is available for the public to view, search and download for research or other purposes. The data is also backed up on a hard drive

and the cloud. The web portal and HarassMap open data will be advertised on our website, and the direct link shared with our contacts and others who approach us with interest in our data.

Keywords

data management plan, sexual harassment, Egypt, crowd sourcing, dmp, research data management

Data Collection

What types of data will you collect, create, link to, acquire and/or record?

We collect different types of data. We will record numeric, audio, PDF, and text, images. The data includes crowdsourced reports that we receive online, reports, comments and messages that we receive on social media, and field data from research projects (interview transcripts, for example).

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

The data will be in different formats, such as XLS, docx, PDF/A, sav, and MP3. These formats are easy to re-use as long as researchers are able to work with the software. In addition, the data will be available and accessible to the user.

What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

At the beginning, we will add the final version for any type of the data, and there will be three other copies of the data, with one copy stored off site (external hard disk), with an access to certain staff members. In addition, the data will be grouped based on the nature of data as follows:

- The data that will be collected from the field, the file's name will include "Field data", with indication for type whether this data is questionnaire or guideline, audio, report, stories, or interview transcription.
- The data that will be collected from our online reporting system, the file's name will include "Map reports".
- The data that will be collected from our social media platform, the file's name will include "socialmedia_ FB" for data from Facebook, and "socialmedia_ TW" for data from twitter. Data from Facebook and Twitter will be structured according to date.

There will also be text descriptions with information about how each dataset is organised.

Documentation and Metadata

What documentation will be needed for the data to be read and interpreted correctly in the future?

There are different types of data, some needs documentation to make the data usable by other researchers and other do not need documentation because it is understood from the title. For the data that required documentation such as field data about sexual harassment, the documentation must include: research methodology used, sample size, variable definitions, assumptions made, format and file type of the data, a description of the data capture and collection methods, explanation of data coding and analysis performed (including syntax files- if available).

The crowdsourced data will also require explanation of how the data was collected and how it can and cannot be used.

How will you make sure that documentation is created or captured consistently throughout your project?

HarassMap will start collecting data retrospectively. For the last five years we have been receiving data but not collecting. For the purpose of this project we will start from collecting the incoming data and at time we will document data collected in the past, moving chronologically. The purpose is to document as much data as we can in the next 6 months – the duration of the project. To make sure that the documentation is created or captured consistently, HarassMap will use part of the funds allocated through this project to recruit 2-3 interns to work on collecting/cleaning data in the data collection phase. They will work closely with HarassMap unit staff from which the data will be collected.

If you are using a metadata standard and/or tools to document and describe your data, please list here.

Currently we are not using any.

Storage and Backup

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

We will store the data on a server with the purpose of long-term storage (years). The exact details on cost and space will be determined once we hire a developer on a consultancy basis for this project.

How and where will your data be stored and backed up during your research project?

The data will be saved on the server, external hard disks and internet cloud . HarassMap does not have a technology expert on-board. Therefore we will use funds available from this grant to hire a consultant who can help us set this system up and maintain it for the duration of the project. After the end of the project, we will allocate the maintenance fee from another grant as soon as possible.

How will the research team and other collaborators access, modify, and contribute data throughout the project?

Each team and collaborators can access it through internet and web application. Each team will have permission to edit the relevant parts .

Preservation

Where will you deposit your data for long-term preservation and access at the end of your research project?

It will be saved on the server and available to the public on the open database/platform (which will be kept online and running as long as we have funds for it).

Indicate how you will ensure your data is preservation ready. Consider preservation- friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

Each project file will be stored on the server on its format and will be linked to its data in the database using some programming algorithms this will store the data on the server as long as we have the server .

We will use the application to manage accounts - each team/section manager can manage his team members remove, update, modify, and create.

Sharing and Re-use

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

We will primarily share raw data that we collected from a research study that we conducted 2011-2013, crowdsourced map reports, and from social media.

Have you considered what type of end-user license to include with your data?

HarassMap has a default copy right policy: <http://harassmap.org/en/copyright/> [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

<http://opendefinition.org/od/2.1/en/>

What steps will be taken to help the research community know that your data exists?

HarassMap will announce this on our website, but we will also keep a permanent icon on the website indicating to users that they can have access to our data. Without announcing it, HarassMap already gets a lot of requests form researchers about sexual harassment.

We will also utilize our contacts with academics in Egypt, UK, and the US to ensure that they know about our open data.

Responsibilities and Resources

Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

Four HarassMap staff members are involved in this project: Director, Marketing and Communications Unit Head, researcher and Admin and HR Manager. Additionally we will use the project budget to offer paid internships to collect some data and to hire a developer as a consultant.

How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

HarassMap has three different staff members to oversee the project and therefore the absence of any of them will not affect the implementation of the project.

**What resources will you require to implement your data management plan?
What do you estimate the overall cost for data management to be?**

Person to collect data from social media: 9000 EGP for interns to document and clean information

Developer: \$2590

Database upkeep (hosting): Will be determined by the end of June 2017.

Ethics and Legal Compliance

If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

The database backend will be accessible only with a user name and password. The data that is accessible for the public, will be checked to ensure anonymity and confidentiality.

Our data is not generally sensitive but sometimes we get reports with names or numbers and in these cases we only publish the reports after we remove this information. The 'sensitive' information will not be available to the public.

We also are not able to share any information that would put HarassMap staff at risk such as reports that defame a person or a place that can file a defamation claim against HarassMap.

How will you manage legal, ethical, and intellectual property issues?

We will prepare an agreement for intellectual property rights that researcher will have to agree to before they have access to the data.