The
HAND-BOOK
OF THE MODERN DEVELOPMENT SPECIALIST
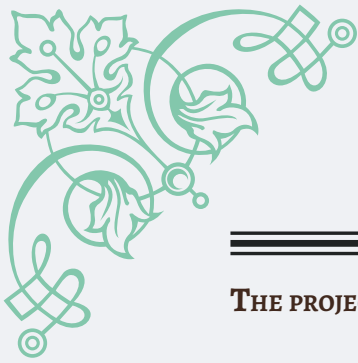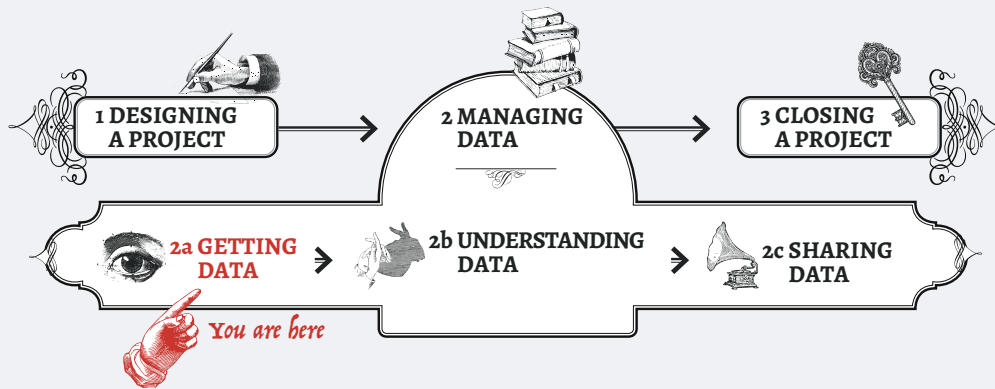
2a

GETTING
DATA

COLLECTING DATA FOR YOUR PROJECT

The project data flow

1 DESIGNING A PROJECT → 2 MANAGING DATA → 3 CLOSING A PROJECT

2a GETTING DATA → 2b UNDERSTANDING DATA → 2c SHARING DATA

You are here

Target audience

Project planners who are getting down to the data details; thinking through what kinds of data the project will need and how to get that data. Issues associated with using datasets from other people are addressed here, as well as things to consider when collecting new data, respecting *data minimisation* principles. Finally, a comprehensive section on *consent* might be of use for those who are wondering what *meaningful consent* actually looks like.

When might this chapter be useful?

When you're getting down into the details of collecting data for your project.

Content summary

Why are you collecting data in the first place? To make sure your data gathering is relevant and actually useful, this section suggests framing your data collection in terms of a concrete question and building out processes and data points based on that specific question. You might not always need to directly collect your own data, either, as there are ways of getting data from existing sources - such as Freedom of Information requests from governments, or simply searching to see if anyone has already published the data you need.

COLLECTING NEW DATA

In the case that you do need to collect your own data, a list of questions is provided to help you plan a responsible data collection process. An exercise is suggested to help you 'define challenge areas' with your team in advance - for this, having a diverse group of people and perspectives will help strengthen the answers.

With responsible data practices in mind, we encourage collecting the minimum data required to answer your question. Practising *data minimisation* can help ensure that resources aren't wasted and that unnecessary (or potentially harmful) data isn't accidentally collected. Broadly speaking, there are a few 'red lines' that you should think carefully about when deciding to collect a new dataset.

LIVE DATA PROCESSES

Systems which collect and process data on an ongoing basis can be the most tricky to plan, as they compress the data cycle, and mean that individual team members might need to play multiple roles. A '*Roles-Responsibilities-Functionality Map*' tool is suggested here to help plan ahead and clearly assign roles and responsibilities within a certain team.

Next, various ways of collecting data are considered, followed by a checklist for tools, security and training that might help think through various technology-related considerations when setting up a data collection process.

WORKING WITH EXISTING DATA

If you can identify existing datasets that support your project activity, there are three primary areas of responsibility to consider. They are explored in more detail here:

### 1. How the data is evaluated for use

Understanding the biases that have gone into the dataset is crucial to making sure it is used in an appropriate and responsible way. Sometimes this is easier said than done, so to help think through these considerations, and gain appropriate contextual information around the data, a short list of questions to go through with your team is included here. You might not need to find the answer to all of them, but actively deciding which of these is most relevant for your project will help in this process.

2. How it is managed

Best practices with regards to data management are explored here, such as keeping multiple, well-labelled versions of your data if it is undergoing transformations. Some legal and licensing considerations are also outlined to make sure you respect the intellectual property restrictions that may (or may not) be associated with the data.

When manipulating the data, it's important to keep in mind *the people in the data* How might various manipulations or combinations of datasets reveal personal or private details about their lives?
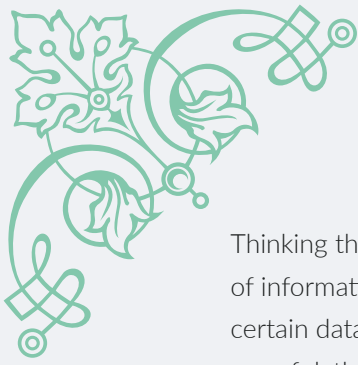
Thinking through carefully what kinds of information could be revealed from a certain dataset is crucial, and can provide a useful, though perhaps dystopian, project planning exercise.

3. How it is presented

Including context to the data is vital, and so is being transparent about what the data does, and doesn't, represent. Being explicit about what is and isn't there can help others interpret and manage your data responsibly.

POWER TO THE PEOPLE

Ultimately, much of the data we are collecting refers to a particular group of people, and it's the people who are reflected in the data who should have the most say over how the data is used. In practice, making sure that this is the case can be trickier than it sounds, though, so here we include a number of considerations to think about to try and ensure that more power lies with those represented in the data, than those doing the collection. With regards to *representation*, for example, there are two main considerations: *inclusion* and *accuracy*.

CONSENT

Different types of consent are discussed in detail in this section. For those who have heard terms like "*informed consent*" used but not really understood what this means in practice or in legal terms, this section may well prove especially useful. Different required elements in a responsible consent procedure are considered here in detail, such as the items needed to be included in a consent agreement, issues that should be brought up with participants in the agreement, and ways of disclosing information to that participant or group of participants.

Developing a strong consent policy within a project or an organisation can help others set up responsible consent procedures more easily, so some components within a responsible consent policy are included here, along with advice on how to implement and maintain such a policy.

Finally, a word of caution is offered: is consent broken? A brief overview of this argument, and some ways to help create meaningful consent, are included here.

Useful resources

Data QualYtI: Do you trust your data?

(an article Hjusein Tjurkmen, Mariyana Hristova, Musala Soft)

http://istabg.org/data-qualyti-do-you-trust-your-data

more at *http://schoolofdata.org/handbook/courses/finding-data/#sthash.8o6YozUJ.dpuf*

Organizations that may be able to help you further on issues you may encounter:

Geeks without Bounds *gwob.org*

Datakind *www.datakind.org*

A checklist for evaluating policies on consent is available at

https://docs.google.com/a/theengineroom.org/document/d/1PJxBAP1rFkjq9p7NuYcN_G5iomfCML-qiMTn5SPPHxE/edit

Solove, Daniel J., Privacy Self-Management and the Consent Dilemma (November 4, 2012). 126 Harvard Law Review 1880 (2013); GWU Legal Studies Research Paper No. 2012-141; GWU Law School Public Law Research Paper No. 2012-141. Available at SSRN: *http://ssrn.com/abstract=2171018*

Launching an SMS code of conduct for Crisis Mapping

http://irevolution.net/2013/02/25/launching-sms-code-of-conduct/