Contents

Acknowledgements

1 Why online research and data visualisation?

Field research among geographically dispersed communities is time-consuming and costly. When people are stigmatised, field research has additional ethical and logistical problems. In many countries lesbian, gay, bisexual and transgender (LGBT) people are both geographically dispersed and stigmatised. Online research methods and tools are therefore particularly interesting instruments for researchers and activists who work with LGBT communities. In countries where same-sex relations are criminal, such as in the Middle East and North Africa region, online communities can be the only way for LGBT people to relate to peers (ILGA 2014). Even in countries where access to social media and publishing on the internet is legally restricted, LGBT people have large online communities (Oosterhoff, Hoang and Quach 2014). This methodology brief outlines the main steps and considerations for choosing research methods and data visualisation among LGBT individuals in resource-poor settings. Although this report focuses on LGBT, online data collection and data visualisation have broader relevance for thinktanks, whose targeted audiences increasingly function in complex digital environments.

Online research is a rapidly changing field driven by its huge market potential. This commercial interest is also one of the dangers to privacy and security of all internet users, with a particular urgency for activists. Security issues are discussed in Section 6. One important aspect of this huge market potential which researchers may have to engage with is gaming. The total number of people around the globe who play games – such as cooking, racing or action games, puzzles or quizzes – on their computer, telephone or iPad, for example, was expected to surpass 1.2 billion by the end of 2013.[1] This is important information for researchers as it affects how people spend their time on the internet and what they expect to find.

Some of the most successful research on sexuality has been through links with commercial dating sites such as OkCupid, the largest free dating site in the United States. For researchers there is a clear interest in using online methodology to reach online communities, but respondents may not be interested in the research. The research questions, methodology and the presentation of results have to take the interests of online respondents into account, in order to generate sufficient attention and responses. For example, although early survey methods such as text-based email surveys are easy and cheap, participation is low because people find them boring. Researchers need to carefully consider whether there are shared areas of interest between them and the respondents. Through OkCupid, for example, researchers were able to obtain data that

Conducting surveys using mobile devices
Source: Image courtesy of kiwanja.net (2014).
Photographer: © Nathan Eagle/MIT (2006), courtesy of Nathan Eagle.

---

[1] See Newzoo *Global Games Market Report 2013*, www.newzoo.com/games-market-data/reports/previous-reports/global-games-market-report/ (accessed 13 August 2014).

provided interesting new intellectual insights on sexuality, while site users received useful practical advice on dating and dating sites.[2] Online preferred reading times tend to be shorter than offline. Data visualisation is helpful here as it can provide clear and powerful pictures of complex social realities.

It is easy for researchers who are looking for methods and tools to work with online communities to get lost among the many available commercial and open source sites. The technology and language can be complicated. In addition, there are many ethical and practical considerations in choosing online research methods and tools, such as technological and language requirements, budget restraints, user characteristics and privacy. Online interviews and focus groups are commonly used, often on Skype™ or by email. While these techniques are relevant, this methodology briefing focuses on open source free tools that have already been tested in resource-poor settings. Because the online world is closely linked to mobile phones, we will also discuss some sites that facilitate mobile data collection, analysis and presentation.

Computer and communication technology are enabling rapid new techniques for data visualisation of massive data sets for public and commercial use. LGBT activists use visualisations to make statements about relevant policies in different parts of the world; for example, in Figure 1.1 the persecution of LGBT people marked in red through to yellow clearly shows a concentration of discriminatory laws in sub-Saharan Africa.

Figure 1.1   Lesbian and gay rights in the world

Source: ILGA (2014).

---

[2] See http://blog.okcupid.com/index.php/about/ (accessed 21 July 2014).

2 Developing a methodological framework: selection of methods and tools

In all the excitement about technological and market opportunities, it is important for researchers to be clear about the exact purpose for the tool they are seeking. Key questions to consider are:

1. **What is the research question that we want to answer?** Developing a clear and specific research question is probably the most important and often the most difficult part of the methodological process. What do we want to know? Why? What are the theoretical assumptions or the hypothesis of the research? What kind of research will be done and where? What – if any – are the specific objectives? Why is this relevant?
2. **What data do we need to answer our research question?** Do we need qualitative, quantitative, or hybrid sets of data to answer our study question? How do we obtain these data (online, offline, both?) Do we want to collect new data or visualise, or could we use data that are already available, for example, on existing websites or fora.
3. **Who is the respondent?** What motivates them? What are the demographic characteristics of our respondents, for example, age, residence, class? If conducting online research, do we know which sites respondents visit and why? How long do they stay on these sites? Is our research question linked to an area of interest to our respondents?
4. **Who are the researchers and what is their technical capacity?** Are we working with researchers who are able, interested and willing to work with online tools? Is the choice of online research driven by respondents' existing expertise, or by expertise or interest in the methodology?
5. **What is the research environment and context?** What is the legal and political online environment of the respondents? Can the site chosen to ask the questions also safely store data? Who owns these data and who controls access? What are the costs and ethics involved?
6. **Who is our target audience for the research?** Is there an overlap between our target audience and the respondents? If these are two separate groups, the demographics and other characteristics of the audience need to be examined in order to develop a communication framework for the research dissemination. Are audiences online or offline? What sources of information do they trust and appreciate?
7. **Language.** What language do your respondents and target audience use? Some tools such as LimeSurvey handle different scripts, such as Vietnamese, while others do not. For use in languages other than English, what you might need is the functionality to ask questions written with the required character set, and to handle answers written with that character set (rather than have the entire application in that character set).

3 Online interviews, face-to-face interviews, or mixed methods?

Whether face-to-face interviews are more or less effective than online interviews depends on many factors (Mann and Stewart 2000). A computer-mediated relationship has well-known limitations, such as a lack of visual clues. These can cause difficulties in assessing how questions and replies should be interpreted, a possible lack of trust on either side, or exclusion of people who are not computer-literate such as elderly or poor rural people. Other problems include a potential lack of privacy and data protection, and lack of understanding of the context in which people respond to questions. However, each of these downsides can also be seen as an advantage. The lack of visual clues can reduce appearance-based bias. People who are less mobile (for example, hospitalised) can be reached more easily. People may also feel more free to share their feelings online than in person. Asynchronous online interviews (using email) can be seen either as 'impersonal', or as enabling people in different time zones to participate. To reduce bias caused by exclusive reliance on one method or the other, much research with LGBT people uses a combination of online and offline methods. Researchers study conversations in online fora or use these to recruit people for face-to-face interviews. For LGBT activists, one of the challenges is to make a transition from invisible online communities to offline 'live' groups of people, who collectively engage in activities such as gay pride parades (Oosterhoff *et al.* 2014).

Using text messages for research, Kenya
Source: Image courtesy of kiwanja.net (2014).
Photographer: © Nathan Eagle/MIT (2006), courtesy of Nathan Eagle.

4 Types of online data collection methods

- Online individual or group interviews[3]
- mobile data collection
- online survey
- poll widgets
- gamification.[4]

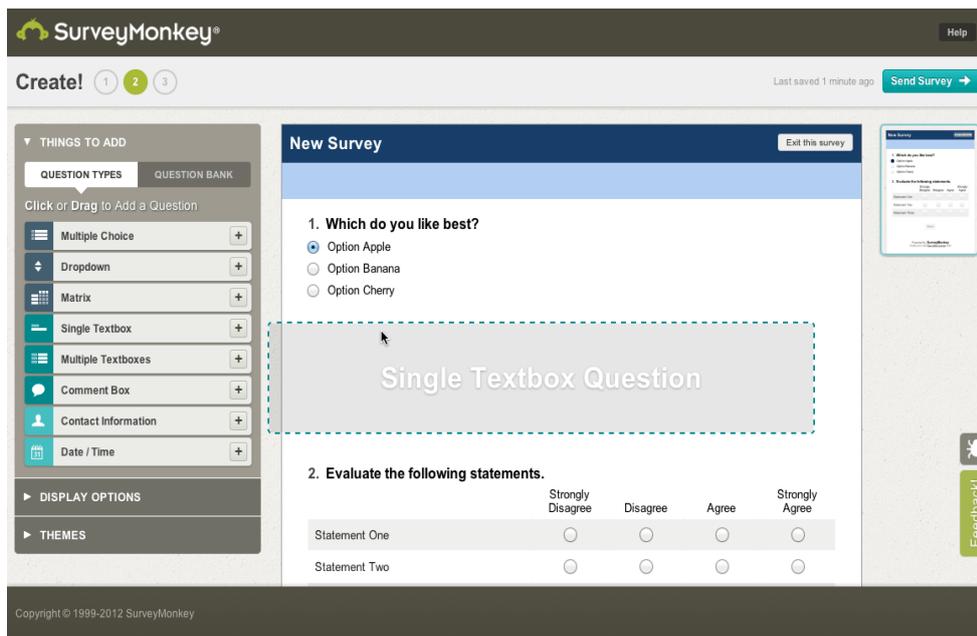| Table 4.1 Mobile data collection	
Open Data Kit (ODK) http://opendatakit.org/	'Open Data Kit (ODK) is a free and open-source set of tools which help organizations author, field, and manage mobile data collection solutions.' ODK provides solutions for users to: • 'build a data collection form or survey • collect the data on a mobile device and send it to a server • aggregate the collected data on a server and extract it in useful formats.' This is achieved through three applications: **ODK Build:** form builder. Surveys can handle decision trees; **ODK Collect:** android app that can handle images, audio, video; **ODK Aggregate:** data storage server. All of the ODK solutions use character encoding called UTF-8, which handles different characters including non-European characters such as Vietnamese.
KoBoToolbox www.kobotoolbox.org/	'KoBoToolbox provides an integrated suite of applications for handheld digital data collection.' Based on ODK but provides offline synchronisation support.
openXdata www.openxdata.org/demo/	openXdata is an open source for data collection; it supports low-cost mobiles and has a visual designer for forms.
formhub https://formhub.org/	formhub is a hosted ODK Aggregate server which can also self-host. Uses ODK applications to interact with the data collected.

---

[3] There are different forms of online synchronous individual and group interviews via common chat technology, such as Skype™ or Google Talk, and asynchronous online interviews (for example, via email). Online interviews can be conducted in a group setting or on a one-to-one basis. These are so common that we do not discuss them further here.
[4] Gamification is the use of game thinking and game mechanics in non-game contexts to engage users in solving problems.

Table 4.2   Online Survey Tools

Survey™ Project www.surveyproject.org/	'Survey™ Project is a free web based survey and (data entry) forms toolkit for processing & gathering data online… written in ASP.NET and C#.'
LimeSurvey www.limesurvey.org/en/	An open source survey application. Drupal™ and WordPress plugins are available to manage survey data.
OpenSurveyPilot http://sourceforge.net/projects/osp/	An open source web-based system for generating complete online surveys and polls.
SurveyMonkey™ www.surveymonkey.com/	Open source web-based surveys that can be created in a few minutes with questions in over 15 formats.

Figure 4.1   SurveyMonkey^{TM} 'create' screenshot

Source: SurveyMonkey™ (2014).

Table 4.3 Poll widgets	
Polldaddy http://polldaddy.com/	Polldaddy enables the user to embed polls in websites. It also integrates with WordPress, Facebook, and other services. Free (limited) accounts; hosted, not open source.
WordPress plugin directory http://wordpress.org/plugins/ search.php?q=poll/	There are various tools and approaches to embed polls in blogs using WordPress and many of the options may be found on this site.

| Table 4.4 Gamification	
www.slideshare.net/hybridlearning/ gamification-tool-and-resources/	This webpage provides resources for games and gamification, including tools and power points

5 Visualisation

Once you have collected your data, visualisation can be an option. Most of the tools listed above record/export data in a simple process, in easily readable formats (such as comma-separated values). The trick with visualising is that it is less about the data itself, but more about the choice of visualisation methods or processes.

The data collected depends on the type of survey questions asked. For example, if you provide a free text box for replies to a question about income, you are more likely to get 'dirty' data[5] than if you allocate a drop-down list of defined income ranges. It may be that even with the best data available, you need to clean some of the data before visualising. There are a number of tools available that enable users to bulk process and clean data, such as Google Refine.

Table 5.1 Visualisation	
Drawing by Numbers https://drawingbynumbers.org/toolsandresources/	This site for activists provides information on why and how to use data visualisation, clean data, and many links to tools and samples of work by other activists.
Introduction to… Data Visualisation www.dmeforpeace.org/sites/default/files/Introduction-to-Data-Visualisation.pdf	This briefing by IDS Knowledge Services provides an introduction to the rationale and steps for data visualisation, as well as sites where useful practical tools can be found.
Many Eyes www-958.ibm.com/software/data/cognos/manyeyes/	Site users can view and discuss visualisations, view and discuss data sets, and create visualisations from existing data.

---

[5] Data which have not yet been checked for common errors such as typing or spelling errors. Using commas instead of full stops in an Excel file, for example, can result in incorrect totals.

Figure 5.1   SurveyMonkey™ analyse and visualise data
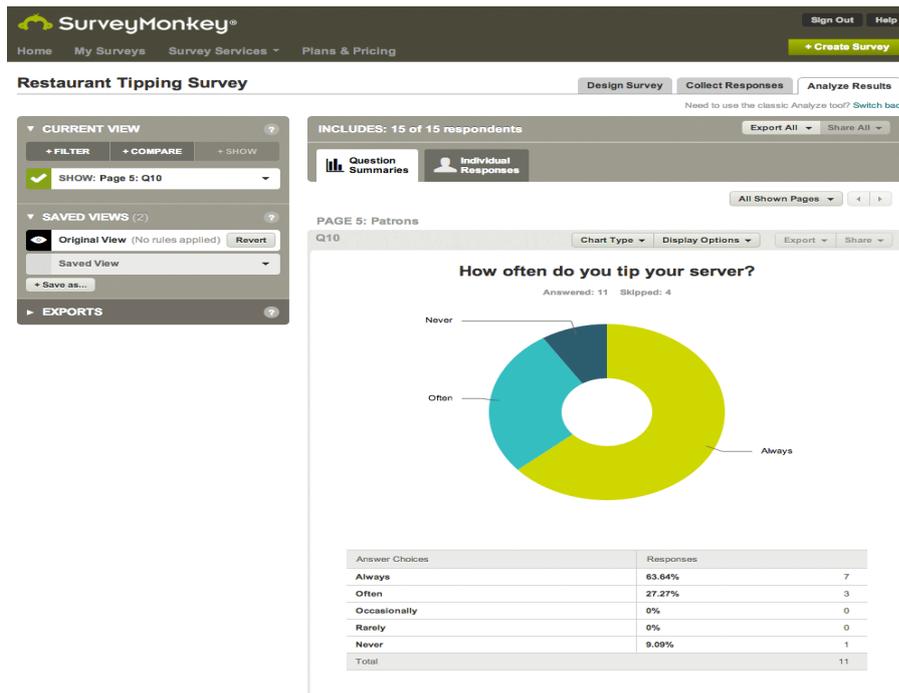
Source: SurveyMonkey™ (2014).

Figure 5.2   OpenStreetMap (OSM)

Source: © OpenStreetMap contributors (OpenStreetMap Foundation 2014).

OpenStreetMap is an online map of the world, created by people who wanted a new, fully adaptable, free-to-use map.

6 Security

Today, data security is a serious issue. The movement of people and personal information can be tracked in multiple ways through data that are collected by mobile phones and social media without the user knowing it. Internet security is an extremely dynamic and complicated field that requires specific expertise. Given that the situation changes constantly, this methodology brief cannot and should not aim to give the reader a definite overview of answers and solutions to common threats.

6.1 Digital security organisations

In order to learn about online security and stay up to date it is probably best to rely on specialists who are working on this issue. There are many non-profit organisations that work on digital rights, internet security and freedom of speech for activists. However, in some countries contacting these organisations in itself holds a certain risk, when one does not know what, and how, data can be traced. Some useful websites and organisations include:

- **Tactical Technology Collective (TCC).** TTC is an international collective that supports activists working on progressive social, environmental and political change to use digital tools safely and effectively with films, toolkits and guides, and by hosting training and events: www.tacticaltech.org
- **Bits of Freedom (BOF).** BOF is 'a digital rights organisation, focusing on privacy and communications freedom in the digital age'. Its website provides a wealth of practical information including security on social media such as Facebook and Twitter, where to find alternative email providers and many links to blogs and events all over the world: www.bof.nl/home/english-bits-of-freedom
- **European Digital Rights (EDRi).** The EDRi website is useful to find the latest information on policy and legislation in the European Union and their (global) implications: https://edri.org
- **Electronic Frontier Foundation (EFF).** Based in the US, EFF is a non-profit organisation 'defending civil liberties in the digital world'. Founded in 1990, it is one of the oldest organisations in the world in this area with a wealth of experience and links: www.eff.org/about.

For activists it is prudent to assume that nothing is 100 per cent secure, whether one looks at data storage or at data collection, analysis or dissemination. However, there are some things that one can keep in mind to help reduce online vulnerability to threats.

6.2 Passwords
- passwords should ideally be long and complex. It is best if they have a combination of upper case and lower case letters, numbers, and symbols such as punctuation marks. A good option can be to make a unique sentence but this is only possible when there is enough space
- passwords should be difficult for outsiders to guess so they should not contain personal characteristics (for example, birthdays, name of dog or children)
- passwords should be secret. Therefore, opinions on storing your passwords online differ greatly. Virus protection such as Norton^{TM} offers a password management facility but it leaves your data in the hands of a company, and in the case of Norton^{TM} you have to renew membership for continued access
- change passwords frequently and use different passwords for different sites. If you use the same password for more than one account anyone who learns that password will gain access to even more of your potentially sensitive information. Some sites are easier to hack than others.

6.3 Social media and digital shadows

Companies such as Facebook are making money out of the data that social media users provide them with for free. Through takeovers, mergers and acquisitions, agreements made by the founders of a company with their users may not be honoured. Many users do not know what data or 'digital shadows' they leave behind on the internet and how it works. Understanding digital shadows is a first step to take control of personal information and improve understanding of how surveillance works. The Me & My Shadow website, created by the Tactical Technology Collective, provides users with a range of free resources to help them learn about their digital shadows, and the traces they leave on the internet: https://myshadow.org.

7 Online data storage and transfer

Most cloud storage services (Dropbox, Google Drive, etc.) allow multiple people to access stored files with different levels of access (i.e. view/edit/delete). Although this might be adequate, it depends on the sensitivity of the data you want to store. Anything you put on the internet (whether behind a passworded account or not) potentially could be accessed, but if you adhere to strong security principles (picking a trusted provider, using strong passwords, changing the passwords regularly) it should be secure. When several people have access to a site, each individual needs to adhere to security principles.

If you were thinking of publishing the data for others to analyse and use there are open data platforms available, such as http://datahub.io/. Obviously any sensitive information would need to be removed before uploading.

When you use online tools to analyse or visualise your data it is wise to look at the conditions of use and privacy as well as (corporate) ownership. Many Eyes (www.manyeyes.com), for instance, offers a wealth of examples and tools for data visualisation. It is sponsored by IBM, whose privacy statement clearly says that they 'collect your personal information for various purposes, such as processing your order, corresponding with you, providing you with a subscription or in connection with a job application'. This may or may not be a problem for you.

If you have sensitive data, encryption is an option, but this can be costly. Online backup service providers build their reputations on keeping your data safe and secure both during file transfer and in the online data storage facilities. To do that they use encryption technology such as Secure Sockets Layer (SSL). Once data files are transferred safely to a cloud, another layer of protection is added to prevent others from accessing these data.

Section 8 applies the process of developing a methodological framework to a case study in Vietnam.

8   Online considerations in practice

Box 8.1   Developing an online research project in practice: transgender people and their livelihood options in Vietnam

Background

The laws in the Socialist Republic of Vietnam promote equality for all citizens. Transgender persons have the right to change their gender, and laws refer to 'persons' rather than 'men' or 'women'. However, the economic rights of transgender persons, such as protection from employment discrimination, are not protected. Transgender persons are faced with severe stigma and discrimination in schools, families and in the workplace. Transsexual people have difficulty with their personal identity papers as they cannot change their identity card, which is an obstacle to obtaining social services, housing and work. Gender roles and norms affect the employment practices, options and preferences of transgender men and women differently. For female transgenders, work in the female-dominated sex industry is a well-documented employment option, in addition to performing as singers, speakers and entertainers at wedding and funeral ceremonies. While transgender men do have difficulties in finding jobs, their situation appears to be better than that of transgender women. Transgender women do more manual and heavy work, including construction work and also can be involved in male-dominated illegal activities such as robbery. Difficulties in negotiating safe sex have resulted in high Human Immunodeficiency Virus (HIV) prevalence rates among transgender people who are now considered a 'key population' in HIV prevention programmes. The data of transgender people are not gender segregated, which makes it impossible to see whether transgender men and women are all equally at risk and affected. Anecdotal evidence suggests that sex work is taboo among male transgender persons.

1. **What is the research question that we want to answer?** We want to explore the different employment options, and preferences of Vietnamese transgender men and women. We assume that gender roles and norms affect their employment options, practices and preferences. This research can help to develop and strengthen gender-specific transgender employment and livelihood opportunities.
2. **What data do we need to answer our research question?** We need hybrid and gender disaggregated sets of data. Quantitative data are needed to gain a picture of what the opportunities are, while qualitative data will help to understand how people interpret this situation, and if and what they would like to change. Online meeting places are the most frequently used channels for transgender persons to find peers. We will collect online survey data, conduct in-depth interviews and group analyses. These will be new data.
3. **Who is the respondent?** We will interview at least 120 transgender men and 120 transgender women from different parts of Vietnam (North, Central and South), preferably of different ages (16–25, 26–35, 36–50 and >50 years old). Because most of the internet users are young, online research has an age and internet access bias. The overall online behaviour and preferences of transgender men and women are unknown. Transgender persons use LGBT fora and social media to gain emotional support and find practical information. There are a number of such LGBT forums. Bangaivn.net attracts both transgender men and women, and is willing to host a survey on the topic of livelihood. Lack of employment opportunities is a major concern of both this group and also the forum hosts. Preliminary investigations online and through informal interviews showed that transgender people would like to know which employers accept them and what kinds of work are available. There is currently no gay-friendly employer nomination mechanism in Vietnam. Invitations to join the research will be distributed through other forums and social media. The survey can also be posted on relevant websites and Facebook. In addition, the survey can be posted on general websites for lesbian and gay people such as bangaivn.net and taoxanh.net.

Box 8.1 (cont'd).

4. **Who are the researchers and what is their technical capacity?** We are working with experienced Vietnamese social science researchers who are able, interested and willing to work with online tools and have experience with SurveyMonkey™. The choice for online research is driven by the fact that the respondents are geographically dispersed and need flexibility in terms of their time. The researchers have no experience with data visualisations. It is unclear how much they know about digital shadows and digital security.

5. **What is the research environment and context?** Vietnamese bloggers who criticise the one-party state have been arrested, but LGBT bloggers and journalists have not been targeted by the state. Although the legal environment is restrictive with regard to freedom of speech and public association with civil society organisations (CSOs), the number and influence of CSOs are growing. Internet usage has grown exponentially since its arrival in Vietnam 15 years ago. An estimated 31 million Vietnamese – one third of the population – use the internet, 70 per cent of whom use Facebook. Sites where research questions are asked cannot and/or are usually unwilling to safely store data. Sites are willing to publish the results of research. It is not clear if and where data can be stored for free. The forums do not have money and members are poor. It is important that preliminary quantitative research results are presented to transgender people for analysis for ethical reasons and to improve the quality of the research. This discussion can take place both in person and online.

6. **Who is our target audience for the research?** The target group for the research overlaps with the respondents as this research aims to help them to gain a better insight into their employment and livelihood options. Other target groups include employers, policymakers and activists. Activists can be reached through online and offline networks. Some employers have openly supported the LGBT cause and can be approached directly through existing networks; others can be informed about transgender persons' capacities and preferences through official and informal networks. Many influential policymakers are offline. Information needs to come from within Vietnam to be trusted and appreciated.

7. **Language.** Vietnamese.

Box 8.2   Methods and tools

Based on the above considerations and discussions with experts from IDS Knowledge Services and the Tactical Technology Collective, which online methods and tools could we use?

We can conduct interviews using SurveyMonkey™ with a forum site, keeping in mind the following with regard to our data:

*   SurveyMonkey™ servers are in the US. If anyone wanted to access the data from our survey legally they would have to file there
*   SurveyMonkey™ is slightly vague about the ownership of data; it belongs to the researcher/users but they also have rights to it in some cases
*   the data collected via SurveyMonkey™ is more valuable/dangerous in aggregate than one individual survey, which is why access to and control of the administrator of the survey's account is key.

We should think of security as having three levels:

Level 1: The administrator of the SurveyMonkey™ survey

The administrator will have an account to set up the survey, administrate, collate results, etc. Some guiding rules are:

- do not spread access to too many people
- keep it to two people maximum
- the administrator (and others) needs to have a good level of security. They should use a https (hypertext transfer protocol secure) server, and a Virtual Private Network (VPN). A VPN will route their internet access through servers elsewhere so that it does not look like someone from Vietnam is trying to access their site
- further information on security for the administrator can be found at securityinabox.org (also in Vietnamese).

Level 2: Referrals (transgender people, friends and fans)

- whoever is emailing links to the survey needs to be discrete about it
- send emails with the link to the site through a VPN
- use language, words and terms that are unique to the transgender community; some language and words only they will know and understand
- develop a questionnaire and fill out the potential answers. Ask yourself, what will happen if all the questions get leaked? Who is at risk and what is at risk? Then ask, what will happen if the answers get leaked? Is there a way to frame questions and answers and demographic information so that people's identities are not revealed?

Level 3: Survey respondents

- for online surveys, there needs to be some offline or secure online awareness-raising to communicate that the information is privileged and private, with a clear statement of what the risks are.

References

ILGA (International Lesbian, Gay, Bisexual, Trans and Intersex Assocation) (2014) *State-Sponsored Homophobia Report 2013: Happiness and Anger,* http://ilga.org/ilga/en/article/o5VlRM41Oq (accessed 15 July 2014)

Kiwanja.net (2014) *Mobile Gallery,* www.kiwanja.net/mobilegallery.htm (accessed 15 July 2014)

Mann, C. and Stewart, F. (2000) *Internet Communication and Qualitative Research: A Handbook for Researching Online,* London: Sage Publications

Oosterhoff, P.; Hoang, T. and Quach, T.T. (2014) *Negotiating Public and Legal Spaces: The Emergence of an LGBT Movement in Vietnam,* IDS Evidence Report 74, Brighton: IDS

OpenStreetMap Foundation (2014) *OpenStreetMap,* www.openstreetmap.org/#map=8/52.270/-1.469 (accessed 5 August 2014)

SurveyMonkey™ (2014) *Brand Assets,* www.surveymonkey.com/mp/brandassets/#usage (accessed 15 July 2014)