

Open Data Management in Agriculture and Nutrition

*This e-learning course is the result of a collaboration between **GODAN Action** partners, including **Wageningen Environmental Research (WUR)**, **AgroKnow**, **AidData**, **the Food and Agriculture Organization of the United Nations (FAO)**, **the Global Forum on Agricultural Research (GFAR)**, and **the Institute of Development Studies (IDS)**, **the Land Portal**, **the Open Data Institute (ODI)** and **the Technical Centre for Agriculture and Rural Cooperation (CTA)**.*



GODAN Action is a three-year project UK's Department for International Development to enable data users, producers and intermediaries to engage effectively with open data and maximise its potential for impact in the agriculture and nutrition sectors. In particular we work to strengthen capacity, to promote common standards and best practice and to improve how we measure impact. [www.godan.info]

UNIT 1: OPEN DATA PRINCIPLES

LESSON 1.2: ETHICS IN OPEN DATA LIFECYCLE



Photo by [Neil Palmer \(CIAT\)](#) licensed under CC BY-SA 2.0

Aims and learning outcomes



The lesson aims to:

- list the stages of the open data lifecycle
- define responsible use of data
- provide an overview on principles of responsible data
- define responsible data concepts
- provide an overview in responsible data challenges in agriculture
- list ethics questions to ask at stages of the open data lifecycle.

After studying this lesson, you should be able to:

- understand stages of the open data lifecycle
- define challenges with use of data
- understand responsible data principles
- become aware of responsible data challenges in agriculture
- identify ethics questions to ask at stages of the open data lifecycle.

Contents

Lesson 1.2: Ethics in open data lifecycle	2
Aims and learning outcomes.....	2
1. Ethical and responsible use of open data	4
2. Responsible data in agriculture	4
3. Principles of responsible data reuse	6
4. Responsible data concepts.....	7
Works Cited.....	9

1. Ethical and responsible use of open data

From its generation to administrative practice, data is involved in different stages which we will refer to here as the data lifecycle. There are various models to identify an open data lifecycle such as that of Van den Broek *et al.* comprising the steps: (1) identification, (2) preparation, (3) publication, (4) reuse, and (5) evaluation (Broek, Veenstra and Folmer). Ethical considerations should be involved in every steps of the open data lifecycle.

Sharing data enables recreation of new insights from existing data but it has also potential to introduce ethical risks. We can't assume that 'open,' 'shared' and 'public' are automatically for the 'public good', particularly when a dataset is a combination of other datasets for which we don't know when the data was first collected or for what purposes. Data has the potential to empower new voices and approaches, but it can also expose the vulnerable and marginalised. Responsible data ethics can often account for the difference between these binaries or polar extremes (Responsible Data Forum).

These challenges place an even bigger responsibility in the hands of intermediaries, who need to be aware of the opportunities and risks of the data that they are working with, and to embed sensitivity and responsibility in their data-handling practices. Adopting a critical approach to avert a 'data for data's sake' methodology will go a long way towards ensuring fairness and balance. Future-proofing is also an issue: what seems unproblematic data right now, for example, may turn out to be very sensitive in the future. Being mindful of the particular vulnerabilities and circumstances of the worst-off within the communities that you work for and with can go a long way towards averting or containing harm.

Engaging with responsible data practices means upholding a certain set of ethical practices with regards to the way you use data to help projects enhance the good they aim to do, and to avoid inadvertent harm (Why Responsible Data).

2. Responsible data in agriculture

Later in this lesson we will see that most of the discourse about ethical use of data relates to the privacy of individuals, often in a medical context. In the

case of the agricultural sector we will see that it may also involve the socio-economic relationships between stakeholders in the value chain.

America's first legal case on agricultural data is given here as a use case in responsible use of data. This was a lawsuit concerning data intermediary AgriStats, which undertakes the anonymisation of farm data. Despite the mission statement of AgriStats being to 'improve the bottom line profitability for our participants by providing accurate and timely comparative data while preserving confidentiality of individual companies', it is believed to have shared individual production data with Tyson Foods, Perdue Farms and other integrators in the value chain. These value-chain partners used data to drive production prices down at the expense of the growers. Data provided by AgriStats allegedly revealed crucial data points of individual farms to the value-chain partners who used this for their benefit. In this case the growers had been sharing data with AgriStats in order to get better comparative data about their efficiency and thus help to improve broiler chicken production. The growers trusted AgriStats that this data would indeed be kept confidential and, although their business model clearly thrives on this trust, in this case data intermediary AgriStats failed to deliver (Schaap).

Actors in the agricultural sector faces responsible data challenges. GODAN's report on Responsible Data in Agriculture (Ferris and Rahman) provides an overview of responsible data challenges focusing on power imbalances. One of the examples given deals with the precision agriculture which provides farmers with information and farm management advice to improve their decision making and to optimise their activity. Most precision farming applications are employed in highly capital-intensive farming systems and most of the access to technologies and data remains in the hands of a few, large-scale farmers and service providers. Specialised companies offer access to software and data to assist with precision agriculture, but only the better resourced companies and farmers can take advantage of these new offerings. The report lists a few prominent challenges in the sector:

- Potential for data breaches. It is difficult to accurately assess what may be the consequences of data breaches from companies or actors who deal with a lot of data within the sector. Data breaches are not uncommon though, and are growing in number; this issue seems certain to be significant in the future. This concern is also common among research institutions that may not yet have developed secure repositories for data.
- Sensitive data. When considering vulnerable communities and contexts, it seems commonly understood within the sector that certain types of agricultural data are sensitive in and of themselves and therefore precautions should be taken in determining whether to collect and share this data at all.

- Data ownership. With increasing amounts of data being created about farming and by farmers, one key issue is around ownership of data. The issues around ownership of data generated through new areas of agriculture technology remain relatively unexplored.
- Vulnerable communities. The issues and tensions mentioned above become even more stark when it comes to particularly vulnerable communities such as indigenous populations, migrant farmers and displaced smallholder farmers who are lacking in basic land rights; women are especially vulnerable in such circumstances.

3. Principles of responsible data reuse

Responsible data has been given a working definition of: 'The duty to ensure people's rights to consent, privacy, security and ownership around the information processes of collection, analysis, storage, presentation and reuse of data, while respecting the values of transparency and openness.' (Responsible Data Forum)

Two basic principles underlie the need for responsible data: empowerment – to empower users to be active participants rather than passive data 'subjects'; and harm avoidance – to ensure that we do no harm and that the way in which we use data and technology does not facilitate or exacerbate harm done by others. Responsible data is not just about technical security and encryption, but also about prioritising the dignity, respect and privacy of the people we work with, and making sure that the people reflected in the data we use are counted and heard, and able to make informed decisions about their lives (Why Responsible Data).

The Responsible Data Forum outlined four underlying reasons to the question of why responsible data reuse:

1. *Legal implications.* Among a larger number of key regulations and directives that are applicable on the EU level is the EU Data Protection Directive, which is currently being transformed into a regulation. The regulation will apply if the data controller or processor (organisation) or the data subject (person) is based in the EU. Furthermore (and unlike the current Directive) the Regulation will also apply to organisations based outside the European Union if they process personal data of EU residents. According to the European Commission 'personal data is any information relating to an individual, whether it relates to his or her private, professional or public life'.
2. *Ethics and integrity.* High ethical standards, respect for dignity and organisational integrity are among the key employee motivators. If an organisation negligently demonstrates a lack of care for the privacy and dignity of others, it either leads to a culture of double standards or may be seen to translate into a lack of care about employees' rights.

3. *Rights and dignity of others.* Having responsible data policies sends a clear signal to all stakeholders that an organisation does in fact care about its affected groups, especially those that are more vulnerable.
4. *Reputation in front of donors, partners and customers.* Having data reuse policies in place sends a clear signal to donors, partners, customers and other stakeholders that an organisation treats its activities with care and high ethical standards. Increasingly the donor community is demanding that such policies are in place if organisations are to receive funding.

Data reusers must do all within their powers to avoid causing any harm to stakeholders that could arise as a direct or indirect result of open data reuse. Many conversations are shaped around a confrontation between the 'do no harm' principle and concepts of transparency and accountability. Although it may appear that this is a confrontation of opposites, the balancing rules for these two principles is that the 'do no harm' applies for the powerless and transparency and accountability is applied to the powerful (Granickas).

4. Responsible data concepts

Returning to the definition of responsible data, the very first concept included is the **right to consent**. Informed consent is the mechanism through which people agree to provide information for research or data collection projects. Generally, consent has been understood as something that is given by individuals during direct interaction with researchers or surveyors, and is composed of three components: disclosure of the research objectives and any risks or negative consequences of participating; the capacity of individuals to understand the implications of participating; and the voluntariness of their participation. It is highly recommended to have the consent policies in written form, both to serve as a mechanism for possible planning and decision-making and also as an indication of high responsibility and ethical standards at organisations.

The largest part of the conversation about responsible and ethical data reuse revolves around the concept of **privacy**. Privacy is concerned with control over information, who can access it, and how it is used. Privacy is generally linked to individuals, families or community groups, and is a concept that is often used to demarcate a line between a 'private' and 'public' sphere. It is useful to have policies in place describing possible risks and their mitigation plan. The key risk, obviously, is to publish data that can uncover particular individuals' private information (Granickas). *Guidelines of Open Data and Privacy* issued by the Government of South Australia listed key risks for privacy as follows:

- causing humiliation, embarrassment or anxiety for the individual; for example, from a release of health data, it might be concluded that an individual accessed treatment for a sensitive sexual health condition;
- impact on the employment or relationships of individuals;

- affecting decisions made about an individual or their ability to access services, such as their ability to obtain insurance;
- resulting in financial loss or detriment;
- posing a risk to safety, such as identifying a victim of violence or a witness to a crime. (Government of South Australia)

Organisations are starting to adopt policies that consider the **security** of their employees in the light of possible dangers resulting from working with information. There is an increased attention paid to organisational policies that adapt security protocols and tactics to encompass: (1) digital information security; (2) physical and operational security; and (3) the psychosocial well-being required for good security implementation (Granickas).

The Open Data Institute (ODI) developed its Data Ethics Canvas to help organisations in identifying and managing potential data ethics considerations including, principles, policies and processes (The Data Ethics Canvas). In a white paper, the ODI explored the relationship between data ethics and legal compliance, some existing data ethics frameworks and ethical considerations in data collection, sharing and use (ODI Ethical Data Handling). The Data Ethics Canvas is recommended for use (ODI the Data Ethics Canvas User Guide) at the start of a project where data collection is likely to impact on individuals or wider society and involving any type of data. The Data Ethics Canvas guides organisations to consider these following potential questions in a project:

- What are your data sources?
- Are there any limitations in your data sources?
- Who has rights over your data sources?
- What policies/ laws shape your use of this data?
- Are you going to be sharing the data with other organisations?
- What is your core purpose for using this data?
- Do people understand your purpose?
- Who could be negatively affected?
- How are you minimising negative impact?
- Who will be positively affected by this project?
- Are you communicating potential risks/issues, if any?
- How can people engage with you?
- When is your next review?
- What are your actions?

These questions can help you identifying the ethical risks in any open data related projects. For more on this topic, please visit the citations of the ODI.

Works Cited

- Broek, T.A. van den, A.F.E. van Veenstra and E.J.A. Folmer. "Walking the extra byte : A lifecycle model for linked open data." 2013. *TNO Innovation for Life*. November 2017.
<<https://repository.tudelft.nl/view/tno/uuid:cbb972b5-4288-48de-b7d2-544f6977eaa3/>>.
- Ferris, Lindsay and Zara Rahman. "Responsible Data in Agriculture." n.d. *GODAN Global Open Data in Agriculture and Nutrition*. November 2017. <<http://www.godan.info/documents/responsible-data-agriculture>>.
- Government of South Australia. "Privacy and Open Data Guideline." 2014. November 2017.
<<https://data.sa.gov.au/sites/default/files/Toolkit/Privacy-and-Open-Data-Guideline.pdf>>.
- Granickas, Karolis. "Ethical and Responsible Use of Open Government Data." February 2015. *European Data Portal*. November 2017.
<https://www.europeandataportal.eu/sites/default/files/2015_ethical_and_responsible_use_of_open_government_data.pdf>.
- "ODI Ethical Data Handling." September 2017. *The ODI*. November 2017.
<<https://www.scribd.com/document/358778144/ODI-Ethical-Data-Handling-2017-09-13>>.
- "ODI the Data Ethics Canvas User Guide." September 2017. *The ODI*. November 2017. <<https://www.scribd.com/document/358778143/ODI-the-Data-Ethics-Canvas-User-Guide-2017-09-13>>.
- Responsible Data Forum. "Hand-Book of the Modern Development Specialist." 2016. *Responsible Data Forum*. The Engine Room. November 2017.
<<https://responsibledata.io/resources/handbook/assets/pdf/responsible-data-handbook.pdf>>.
- . "Responsible data is." 2014. *Responsible Data Forum*. November 2017.
<<https://responsibledata.io/about/>>.
- Schaap, Ben. "Why anonymisation of farm data is important when sharing data." 4 May 2017. *GODAN Global Open Data for Agriculture and Nutrition*. November 2017. <<http://www.godan.info/blog-posts/why-anonymisation-farm-data-important-when-sharing-data>>.
- "The Data Ethics Canvas." August 2017. *The ODI*. November 2017.
<<https://theodi.org/the-data-ethics-canvas>>.
- "Why Responsible Data." 2016. *Responsible Data Forum*. November 2017.
<<https://responsibledata.io/resources/handbook/chapters/chapter-00-introduction.html#why-responsible-data>>.