

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 1.1: WHAT IS OPEN DATA



Photo by [Vostoc 91](#) licensed under CC BY-SA 2.0

Aims and learning outcomes



The lesson aims to:

- introduce open data and its principles
- identify benefits and challenges of open data for the agricultural sector
- provide an overview about key actors involved in the open data publishing
- present use cases in agriculture related to benefits of open data.

After studying this lesson, you should be able to:

- acquire knowledge about definition of open data and its principles
- recognize the value of open data in agriculture
- define challenges of moving to an open data landscape
- understand the benefits of publishing and using open data.

Contents

Lesson 1.1: What is open data.....	2
Aims and learning outcomes.....	2
List of figures.....	4
1. Introduction to open data	5
2. What is open data?	6
3. Open data principles	8
4. Benefits of open data	9
5. Challenges	11
Works Cited.....	15

List of figures

Figure 1 The Data Spectrum by the ODI licensed under CC BY	7
Figure 2 The data lifecycle by Mushonz licensed under CC BY-SA 4.0 via Wikimedia Commons	10
Figure 3 Data overview by GODAN (Responsible Data in Agriculture)	14

1. Introduction to open data

In today's world, it is possible to forecast the future much better than ever or to answer seemingly complicated questions much more quickly based on data. Such questions might be: Where does our food come from? Can we manage risks in our farm and take control measures against droughts or pests? Are we able to predict problems such as floods or low yields? Can we make informed decisions on what to grow, what treatment to apply, when to plant, treat or harvest? Technologies today allow us to build services to answer these questions but data only offers these opportunities when it is usable.

The notion of open data has been around for some years. The term 'open data' was first used in 1995 and the concept of open public data was defined in 2007 in a meeting of leading internet thinkers and activists including Tim O'Reilly and Lawrence Lessig (Chignard). Considerable amounts of data today are generated by the public sector, e.g. soil surveys, cultivar registrations, pesticide residues, health care, defence industries, infrastructure, public education, and telecommunications. In 2009 various governments, such as USA, UK and Canada, launched open government initiatives to open up their public information.

Open access to research and open publication of data are vital resources for food security and nutrition, driven by farmers, farmer organisations, researchers, extension experts, policy makers, governments, and other private-sector and civil-society stakeholders participating in 'innovation systems' and along value chains. Lack of institutional, national and international policies and openness of data limit the effectiveness of agricultural and nutritional data from research and innovation. Making open data work for agriculture and nutrition requires a shared agenda to increase the supply, quality, and interoperability of data, alongside action to build capacity for the use of data by all stakeholders (GODAN).

The agriculture sector is creating increasing amounts of data, from many different sources. From tractors equipped with GPS tracking, to open data released by government ministries, data is becoming ever more valuable, as agricultural business development and global food policy decisions are being made based upon data. But the sector is also home to severe resource inequality. The largest agricultural companies make billions of dollars per year, in contrast to subsistence farmers growing just enough to feed themselves, or smallholder farmers who grow enough to sell on a year-by-year basis. (Ferris and Rahman)

When it comes to data and technology, these differences in resources translate to stark power imbalances in data access and use. Farmers transfer data to companies despite it remaining unclear how that data will be used. The most well-resourced actors are able to delve into new technologies and make the most of those insights, whereas others are unable to take any such

risks or divert any of their limited resources. Access to and use of data has radically changed the business models and behaviour of some of those well-resourced actors, but in contrast, those with fewer resources are receiving the same, limited access to information that they always have (Ferris and Rahman, Responsible Data in Agriculture).

2. What is open data?

The open data movement has been advocated strongly by governments to allow others to benefit from their data and their desire to be transparent, but research institutions and the private sector also generate data which they are willing to share as a common good (Gray).

Open data is *data that can be freely used, reused (modified) and redistributed (shared) by anyone* (Open Knowledge International). The Open Data Handbook emphasizes the importance of the definition of open and highlights key features about open data:

Availability and Access: The data must be available as a whole, and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form. Managing data can be costly in terms of time and resources needed. An example of costing for data management can be seen at UK Data Service (UK Data Service).

Reuse and Redistribution: The data must be provided under terms that permit reuse and redistribution including intermixing with other datasets.

Universal Participation: Everyone must be able to use, reuse and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed (Open Knowledge International).

Data exists on a spectrum. It can be closed, shared or open. Dataset may include sensitive information for security, personal or commercial reasons. For instance, health records may cover sensitive data which raise privacy issues. For these reasons, data can be closed or shared to limited persons or groups but not licensed to permit anyone to access, use and share it. Whether big, medium or small, whether state, commercial or personal, the important thing about data is how it is licensed. For data to be considered open, it must be:

- accessible, which usually means published on the web
- available in a machine-readable format
- with a licence that permits anyone to access, use and share it – commercially and non-commercially.

The Data Spectrum in Figure 1 developed by The Open Data Institute (ODI), illustrates the degree of openness of data and helps to understand the language of data (The ODI).

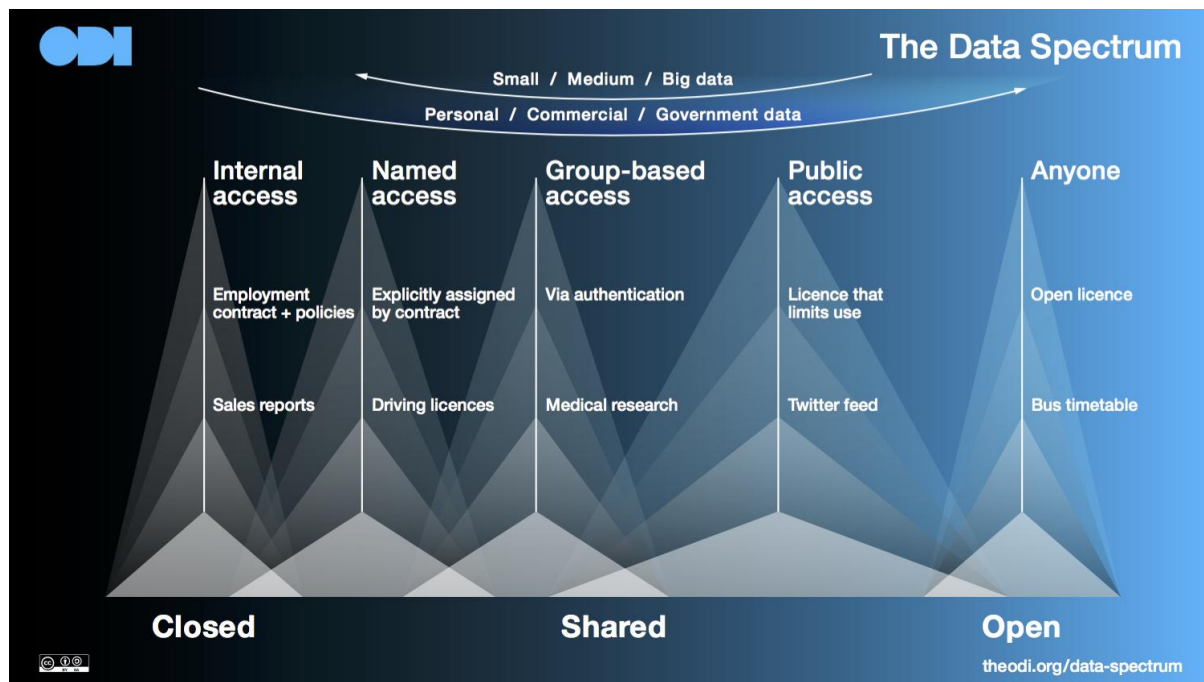


Figure 1 The Data Spectrum by the ODI licensed under CC BY

Many individuals and organisations collect a broad range of different types of data in order to perform their tasks. Government is particularly significant in this respect, both because of the quantity and centrality of the data it collects, but also because most of that government data is public data by law, and therefore could be made open and available for others to use (Open Knowledge International).

There are many kinds of open data that have potential uses and applications:

- **Culture:** data about cultural works and artefacts – for example titles and authors- and generally collected and held by galleries, libraries, archives and museums
- **Science:** data that is produced as part of scientific research, from astronomy to zoology
- **Finance:** data such as government accounts (expenditure and revenue) and information on financial markets (stocks, shares, bonds etc.)
- **Statistics:** data produced by statistical offices such as the census and key socioeconomic indicators
- **Weather:** many types of information used to understand and predict the weather and climate
- **Environment:** information related to the natural environment such the presence and level of pollutants, the quality of rivers and seas (Open Knowledge International).

3. Open data principles

[The Open Definition](#) makes precise the meaning of 'open' with respect to knowledge, promoting a robust common in which anyone may participate and interoperability is maximised. Knowledge is open if anyone is free to access, use, modify and share it – subject at most to measures that preserve provenance and openness (Open Definition 2.1).

Open data must comply with an open licence or a status. It must be in a public domain or under an open licence. Without a licence, the data can't be reused.

It must be accessible and downloadable via the internet. Any additional information necessary for licence compliance must also accompany the work, such as an attribution to say that people who use the data must credit whoever is publishing it or a share-alike requirement to say that people who mix the data with other data have to also release the results as open data.

Open data must be in a machine-readable form which is processable by a computer and where the individual elements of the work can be easily accessed and modified. It must also be in an open format which places no restrictions, monetary or otherwise, upon its use and can be fully processed with at least one free/libre/open-source software tool.

The licence used for the open data should be compatible with other open licences. It should permit free use, redistribution, creation of derivatives, and compilation of the licensed work. It must allow any part of the work to be freely used, distributed, or modified separately from any other part of the work or from any collection of works in which it was originally distributed. The licence must not discriminate against any person or group.

The Open Data Charter, which is a collaboration between over 70 governments, agrees on six principles for how governments should be publishing information. Each of them is explained below briefly. On their site, the Charter also provides detailed action items to achieve each of these principles (Open Data Charter).

Open by Default: Free access to and use of government data (data held by national, regional, local, and city governments, international governmental bodies, and other types of institutions in the wider public sector) brings a significant value to society and the economy, and the government data should, therefore, be open by default. Resources, standards, and policies for the creation, use, exchange, and harmonisation of open data should be globally developed, adopted and promoted as long as citizens are confident that open data will not compromise their right to privacy.

Timely and Comprehensive: Data may require time, human and technical resources to be released and published. It is important to identify which data to prioritize for release by consulting with the data users. The data must be comprehensive, accurate, and of high quality.

Accessible and Usable: Opening up data enables stakeholders to make informed decisions. The data should be easily discoverable and accessible, and made available without any barriers.

Comparable and Interoperable: The data should be published in structured and standardised formats to support interoperability, traceability and reuse. It should also be easy to compare within and between sectors, across geographic locations, and over time in order to be the most effective and useful.

For Improved Governance and Citizen Engagement: Open data strengthens governance, and provides a transparent and accountable foundation to improve decision-making. It enables civic participation and better-informed engagement between governments and citizens.

For Inclusive Development and Innovation: Openness stimulates creativity and innovation. Open data by its nature offers an equitable resource for all people regardless of where they come from or who they are and provides a less digitally divided environment to access and use the data.

4. Benefits of open data

The benefits of open data are diverse and range from improved efficiency of public administrations, economic growth in the private sector to wider social welfare.

Performance can be enhanced by open data and contribute to improving the efficiency of public services. Greater efficiency in processes and delivery of public services can be achieved thanks to cross-sector sharing of data, which can for example provide an overview of unnecessary spending.

The economy can benefit from an easier access to information, content and knowledge in turn contributing to the development of innovative services and the creation of new business models.

Social welfare can be improved as society benefits from information that is more transparent and accessible. Open Data enhances collaboration, participation and social innovation (European Data Portal).

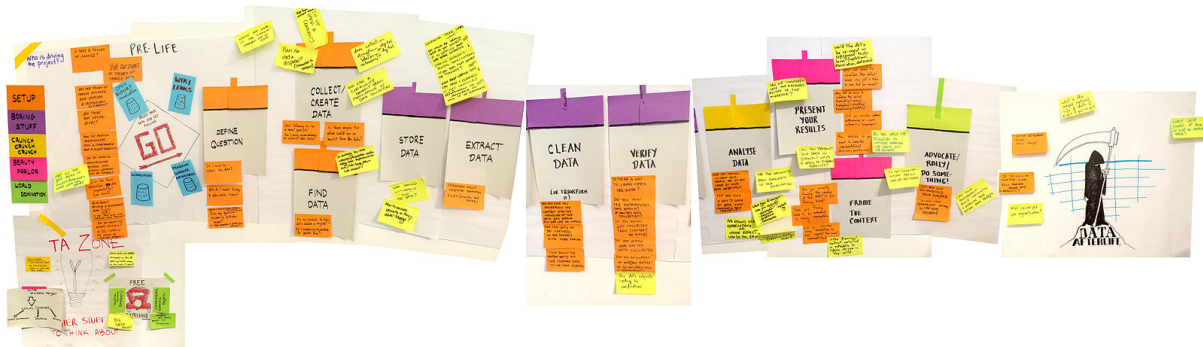


Figure 2 The data lifecycle by Mushonz licensed under CC BY-SA 4.0 via Wikimedia Commons

The availability of detailed open data is essential to improve delivery of services at the local level. Such cases are mySociety [https://www.mysociety.org], the Hungarian 'right to know' portal, and Fix my Street [http://www.fiksgatami.no] Norway. To support the emergence of new data-driven businesses and the growth of existing ones, governments need to publish key datasets. By growing economies and improving services, open data allows governments to make savings in key areas, like provision of healthcare, education and utilities. In the UK, open data helped reveal £200 million of savings in the health service. In France, energy data is being used to drive more efficient energy generation practices (The ODI).

GODAN's report on 'How can we improve agriculture, food and nutrition with open data?' specifies three ways that open data can help solve practical problems in the agriculture and nutrition sectors.

- *Enabling more efficient and effective decision making.* Open data enables computers to pull data from various sources and process it for us, and does not rely on humans to interpret and integrate information contained in web pages. Open data underpins new products and services by presenting information from a wide range of sources that helps everyone from policy-makers to smallholders find gaps in markets or fine-tune their products or services. A good example of it is in fisheries. In South Africa, the fisher community in collaboration with the University of Cape Town co-designed a suite of apps to support and improve the small-scale fisheries industry. Abalobi Fisher is one of those apps which provides valuable information about the weather and climate from open sources, plus records data about fisher practice and catch information. The hope is that it will showcase small-scale fisheries as a vital and valuable resource as well as a legitimised livelihood, not just to the local communities, but to the country as a whole. A relevant use case can be watched at Open Water – GODAN Documentary Web Series at <https://youtu.be/sxY-skVv7Mk>.
- *Fostering innovation that everyone can benefit from.* As a raw material for creating tools, services, insights and applications, open data makes it inexpensive and easy to create new innovations. When data is open for all to experiment with, there is no need to invest large amounts in

repeating already completed trials. When data is openly licensed, it also allows for novel combinations with other data to gain new insights. A story of Andrew from Allington, UK, who works alongside his family on their arable and dairy farm is a good example of using data to apply precision farming tactics to their land. The family has taken advantage of new tools and technology that allow them to easily view satellite data of their land, which has been opened up by the European Space Agency. His story can be watched at Open Skies – GODAN Documentary Web Series at <https://youtu.be/MJ0VX7CqmSc>

- *Driving organisational and sector change through transparency.* Transparency around targets, subsidy distribution and pricing, for example, creates incentives which affect the behaviour of producers, regulators and consumers. By requiring companies, government departments and other organisations to publish key datasets – performance data, spend data or supply-chain data, for example – governments, regulators and companies can monitor, analyse and respond to trends in that sector. More importantly, publishing this data across a sector can ultimately transform how products and services are delivered. We can refer here to the same example, Abolobi Fisher from Open Water in the first section, of using data about fisher practice for small-scale fisheries industry to make informed decisions.

Providing farmers with more accurate, accessible, timely information – from large agriculture groups to the individual smallholders – will help to ensure food commodity markets function well in future. Progress will be driven largely by providing better access to accurate, timely information for individual smallholder farmers, businesses and policy-makers alike. Open data can and should be part of the solution. Open data promotes transparency across the sector to accelerate progress, identify areas for improvement and help create new insights (Carolan, Smith and Protonotarios) Agmarknet [<http://agmarknet.gov.in>] in India is a good example of providing market information with more than 2700 data sources to the farmers, traders, policy makers and other stakeholders for better production and marketing decisions. The rice producer's federation of Colombia keeps data sets historically and helps small and medium growth farmers by measuring climate, yields and farming practices related to rice-growing in the country. You can watch the story of Blanca, who runs her farm outside the town of Ibagué with the assistance of early-warning systems and weather data, at Open Climate – GODAN Documentary Web Series at <https://youtu.be/y9SwlpHT5A4>.

5. Challenges

Open data acts as change agent. Implementing an open data initiative often involves cultural and institutional change. Opening data goes far beyond putting data on a website under an open licence. Applying the technology is relatively easy when compared with bringing about a cultural change, which

can be much harder (European Commission). It requires consulting with potential data users internally within an institution as well as the external stakeholders.

However, this difficulty of adopting a change does not stop the amount of data which is increasingly becoming openly available. There are still challenges related to data management, licensing, interoperability and exploitation. There is a need to evolve policies, practices and ethics around closed, shared, and open data. In response to these challenges at the policy and technical levels, you will find detailed instructions in the following units and the lessons. Rather an unusual categorisation of the open data challenges was made by The Open Data Institute (ODI). These challenges (Tennison) are identified in the following paragraphs and considerations of how to overcome the challenges are briefly included in each item.

Free is not Always Open. Making data freely available is only the first step in making it open for all to reuse. There are layers of barriers between information that anyone can get and data that anyone can use. This is because making the data open is not always about licences or the format issues, but it is also about comprehension and access. Open data should not require additional time, resources and expertise to be used.

Open is not Always Free. This means that open data should not be assumed to be free of charge. High standards in open data releases and lowering the barriers for data users can be costly to the data publishers. It is known that open data business models exist to help public sector and third-sector organisations such as associations and social enterprises save money and deliver or make money by publishing open data (Tennison, Open Data Business Models), (Zeleti, Ojo and Curry).

Analysis is not Always Easy. The diversity of tools available can make open data easy to visualise. However easily made analyses should not stop us criticising the figures. Such analyses should be carefully treated whether they represent comprehensive and accurate figures or not. This matter when people make decisions based on analyses.

Open Data is not Always 'Good'. Open data helps people to make better informed decisions, but it can equally mislead people into making poor decisions, or enable individuals to make such decisions which lead to a more divided society. However, at least open data provides equitable opportunity for everyone to access the data and criticise. Privacy rights, completeness, accuracy, reliability and relevance should be considered in data release and publish.

These challenges are provided here in broad terms that apply across every sector. When it comes to agriculture and nutrition in particular, they no doubt face these same challenges. However, these challenges are best addressed

at the level of a particular problem in a specific field, where standards can be identified or developed, and data released as part of solving a problem. This is especially true when advocates can point to a clear theory of change. GODAN addresses this issue with care and sets out five strategic steps (Carolan , Smith and Protonotarios) for pursuing solution-focused open data initiatives for agriculture and nutrition:

- Engage with the growing open data community, including key problem owners and experts at GODAN, to identify the challenges that open data can help solve.
- Build open data strategies and projects with a focus on finding solutions to agriculture and nutrition problems.
- Develop the infrastructure, assets and capacities for open data in relevant organisations and networks.
- Use open data and support users of relevant data.
- Learn through ongoing evaluation, reflection and sharing to ensure we can all continue to improve our practice.

6. Open data actors

Agricultural data ecosystem involves many private and public-sector actors collecting, analysing and using different bits of data to inform their actions internally, and sometimes also externally with other stakeholders. In the context of this course, it is important to identify the players in the field and see how their role is related to open data.

Public sector actors, such as agricultural, economic and statistical agencies, collect, aggregate and share relevant data within this sector. This data may be simply collected and held, but also may be shared with other government agencies to assist in policy-making decisions or opened and disseminated widely. Some of the key datasets collected include production yields, livestock, weather, market prices and farmers registries.

Researchers from universities, think tanks, institutes, organisations and companies collect and analyse data on subjects from plant sciences, to animal sciences, to soil or climate and many more. Others collect data from farmers via surveys or interviews to understand local markets and farmer constraints. Institutions range from ultra-specialised, commercially focused ones to large, international organisations working on global issues, such as conservation and food security. Bigger agricultural companies often have dedicated Research & Development departments.

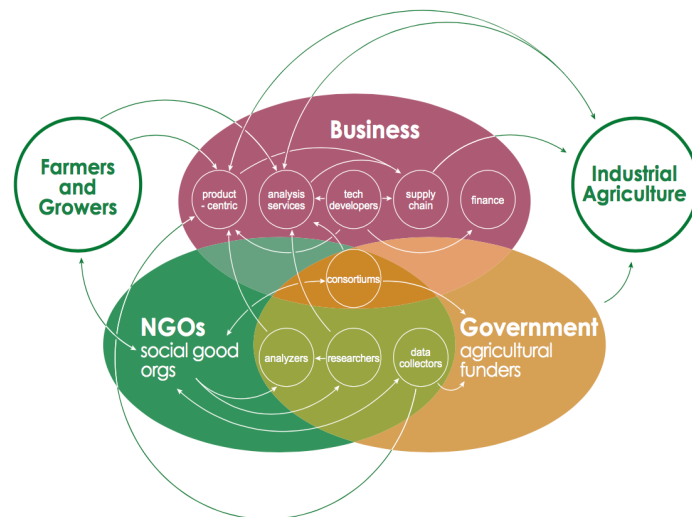


Figure 3 Data overview by GODAN (Responsible Data in Agriculture)

Agribusinesses collect, analyse and use data to inform changes to the services or products they intend to market to clients. Data may be aggregated from a variety of sources, be they in-house, from government agencies or from their clients themselves. Businesses market products, such as agricultural equipment, fertiliser and seeds, or services, such as satellite imagery and financial services.

Farmers produce primary-source agriculture data on their own farms. It may be collected by the previously identified entities or, in large-scale enterprises, analysed in-house. Farmers may use the information produced from this data, or from outside sources, such as the public sector, service providers or research institutions, to inform their farming practices. According to the interviews with farmers held by GODAN (Global Open Data for Agriculture & Nutrition), the most valuable information to growers includes data on weather, soil and land, property ownership and markets (Ferris and Rahman, Responsible Data in Agriculture).

WORKS CITED

- Carolan , Liz, et al. “How can we improve agriculture, food and nutrition with open data?” 14 December 2015. *GODAN Global Open Data in Agriculture and Nutrition*. November 2017. <<http://www.godan.info/documents/how-can-we-improve-agriculture-food-and-nutrition-open-data>>.
- Chignard, Simon . “A brief history of Open Data.” 29 March 2013. *Paris Innovation Review*. <<http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data>>.
- European Commission. “Open data: agent of change.” n.d. *European Data Portal*. November 2017. <<https://www.europeandataportal.eu/elearning/en/module3/#/id/co-01>>.
- European Data Portal. *Benefits of Open Data*. n.d. November 2017. <<https://www.europeandataportal.eu/en/using-data/benefits-of-open-data>>.
- Ferris, Lindsay and Zara Rahman. “Responsible Data in Agriculture.” 2016. *The Engine Room Library*. 2017. <<https://library.theengineroom.org/agriculture/>>.
- 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>>.
- GODAN. *Statement of Purpose*. n.d. November 2017. <<http://www.godan.info/pages/statement-purpose>>.
- Gray, Jonathan. “Towards a Genealogy of Open Data.” September 2014. *General Conference of the European Consortium for Political Research in Glasgow 3-6 September 2014*. November 2017. <<http://dx.doi.org/10.2139/ssrn.2605828>>.
- Open Data Charter. *Principles*. September 2015. November 2017. <<https://opendatacharter.net/principles/>>.
- Open Definition 2.1*. n.d. <<http://opendefinition.org/od/2.1/en/>>.
- Open Knowledge International. “What is Open Data?” n.d. *Open Data Handbook*. November 2017. <<http://opendatahandbook.org/guide/en/what-is-open-data/>>.
- . *What is Open?* n.d. November 2017. <<https://okfn.org/opendata/>>.
- . *Why Open Data?* n.d. November 2017. <<https://okfn.org/opendata/why-open-data/>>.
- Tennison, Jeni. *Four Challenges for Open Data*. 18 September 2012. November 2017. <<https://theodi.org/blog/four-challenges-open-data>>.
- . *Open Data Business Models*. 20 August 2012. November 2017. <<http://www.jenitennison.com/2012/08/20/open-data-business-models.html>>.
- The ODI. *Open Data In a Day*. n.d. November 2017. <<http://www.fiksgatami.no>>.
- . *The Data Spectrum helps you understand the language of data*. n.d. November 2017. <<https://theodi.org/data-spectrum>>.
- UK Data Service. *Costing Data Management*. n.d. November 2017. <<https://www.ukdataservice.ac.uk/manage-data/plan/costing>>.
- Zeleti, Fatemeh Ahmadi, Adegboyega Ojo and Edward Curry. “Emerging Business Models for the Open Data Industry: Characterization and Analysis.” 18 June 2014. *Proceedings of the 15th Annual International Conference on Digital Government Research*. Aguascalientes. November 2017.