Module 1 •

What is Open Data?

Authored:
Javiera Atenas •

Contributors:
Dra. Carla Bonina, University of Surrey, UK •
Dr. Juan Pane, University of Asuncion, Uruguay •
Juan Belbis, Universidad Austral, Argentina •

This project, Understanding data: praxis and politics is funded by the EPSRC and the Human Data Interaction Network+, under the grant number EP/R045178/1. **Caroline Kuhn H.** is the Principal Investigator and responsible for the project. •

Atenas, J., Bonina, C., Pane, J. & Belbis, J. (2021). What is open data?. In C. Kühn, J. Atenas & L. Havemann (Eds.), Understanding Data: Praxis and politics. HDI - Data, Praxis and Politics. https://doi.org/10.5281/zenodo.4783601 •

• Abstract and description of the module •

This module is focused on developing a basic understanding of data and how it permeates different aspects of society. This involves knowledge of how it is structured, the politics behind this and implications regarding transparency. We understand data as a set of values of qualitative or quantitative variables about one or more persons or objects. It is transformed into information when data is created, extracted, elaborated and used with pre-established objectives. The information system, made up of data of the same or different form (the data set is defined as a "dataset"), is transformed into knowledge when it is interpreted through tools, applications, methods, indicators, etc.

The first unit in this module presents an overview of data and introduces the concepts of open data, showcasing different techno-political elements of its structure including principles and standard for publication. The second unit presents an introduction to open science and open science data, and furthermore, principles for scientific data management. Finally, the third unit showcases the importance of fostering innovation with open data to promote the creation of sustainable business models.

• Learning outcomes •

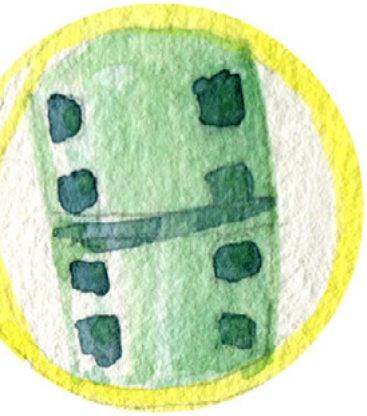
1. Understanding the basics of data and open data

2. Understanding the key principles of open data

3. Understanding the key concepts of data management

4. Acquire skills to publish data in open formats

5. Develop ideas to innovate using open data



• Introductory multimedia • video-podcast •

Nigel Shadbolt “The Value of Openness - The Open Data Institute and Publically Funded Open Data” https://www.youtube.com/watch?v=8wu5B0eWBsk

• Glossary of terms and acronyms • From wikipedia when possible •

Data: is a symbolic representation that describes facts, conditions, values or situations. Data

Information: is an organised set of processed and related data in a way that allows us to communicate or acquire knowledge. Information

Open data: are data that can be freely used, reused and redistributed by anyone, subject only, at most, to attribution and distribution requirements with the same licence link. Open data

Open data use licence: is the authorisation to use the data issued by the source that owns the copyright of the data.

Dataset: is a collection of organised data records where each element has the same structure, ordered for processing by a computer.

Open data catalogue: is an information management system that aims to be a single point of reference for those who want to search for and access data. It is made up of a management system for datasets and their metadata, which provides users with tools to speed up the publication, access, search and navigation of data. Open data catalogue

Co-creation: those processes or activities where at least two actors (for example, public, private, governmental or civic) collaborate in the realisation of a project to achieve a certain result.

Hackathon: a marathon of ideas, design and software prototyping, which brings together specialists, technical programmers, designers and entrepreneurs to work collaboratively. The days tend to last between 24 and 48 hours and usually have specific challenges to guide the event. Hackathon

Digital innovation platform: offers functionalities on which independent companies, developers or innovators can build complementary products, services or technologies.

• Recommended reading •

1. Johnson, J. A. (2014) 'From open data to information justice', Ethics and Information Technology, vol. 16, no. 4, pp. 263–274. Doi: 10.1007/s10676-014-9351-8

2. Baack, S. (2015) 'Datafication and empowerment: how the open data movement re-articulates notions of democracy, participation, and journalism', Big Data and Society, vol. 2, no. 2, pp. 1–11. Doi: 10.1177/2053951715594634

3. Arzberger, P., et al., (2004) 'Promoting access to public research data for scientific, economic, and social development', Data Science Journal, vol. 3 (November), pp. 135–152. Doi: 10.2481/dsj.3.135

4. Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P. L., Görögh, E., Helbig, K., ... & Ross-Hellauer, T. (2018). The open science training handbook https://www.fosteropenscience.eu/content/open-science-training-handbook

5. Davies, T., Walker, S. B., Rubinstein, M., & Perini, F. (2019). The state of open data: Histories and horizons (p. 592). African Minds. https://stateofopendata.od4d.net/

6. Gurin, J., Bonina, C., & Verhulst, S. (2019) Open Data Stakeholders - Private Sector. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre. Print version doi: 10.5281/zenodo.2677777

• Key Complementary resources •

1. Open Standards for Data
https://standards.theodi.org/#:~:text=Open%20standards%20for%20data%20are,adopt%20open%20standards%20for%20data.

2. Tennant, J. (2020). A value proposition for Open Science.
https://osf.io/preprints/socarxiv/k9qhv/

3. Train-the-trainer card game for Open Science training
FOSTER https://www.fosteropenscience.eu/content/train-trainer-card-game-open-science-training

4. Open Data Innovation Week tools
https://labs.webfoundation.org/projects-2/open-data-innovation-week-2/

1 • 1 • Introduction to data •

Contributors: Juan Ignacio Belbis and Juan Pane

We need to look at the whole society and think, "Are we actually thinking about what we're doing as we go forward, and are we preserving the really important values that we have in society? Are we keeping it democratic, and open, and so on?"

Tim Berners Lee

• Introduction •

Data are characteristics or information, usually numerical, that are collected through observation. In a more technical sense, they are a set of values of qualitative or quantitative variables about one or more persons or objects, while a datum is a single value of a single variable. Data are transformed into information when they are created, extracted, elaborated and used with pre-established objectives. The information system often made up of data of the same or different type (the data set is defined as a "dataset"), is transformed into knowledge when it is interpreted thanks to tools, applications, methods, indicators, etc.

Data can be small or big, private, personal, governmental, military, scientific, public, confidential, commercial, financial or open, and normally pertain to information delivered in machine-readable file formats (machine-readable) in a format known as raw data. The most common formats are integer, floating-point number, character, string and Boolean. With the constant evolution of technology, the informative content and the data held by public administrations represent excellent opportunities to promote transparency in the actions of governments and administrations. Moreover, they can offer more efficient services and, since they facilitate reuse by other public and private subjects, they also can be used in areas other than those for which they have been produced or collected.

Knowledge, in practice, acquires the value of awareness - in the case of open data these can be defined as "collective", understood as being for the "common good" - when used for change and the improvement of reality (the facts).

Whilst data are features of information that are collected through observation, information is understood as a symbolic representation that describes facts, conditions, values or situations, collected and arranged in an appropriate way to fulfil the objective of the institution that manages it. On their own, these values lack a semantic value, that is, they do not have a meaning for someone, so they do not add value to the recipient of the message. For these data to make sense, they must be processed, associated or grouped within the same context to form information. Thus, we can conclude that information is an organised set of processed and related data in a way that allows us to communicate or acquire knowledge.

1 • 1.1 • Understanding open data •

According to the International Open Data Charter, "Open Data is digital data that is made available with the technical and legal characteristics necessary so that it can be freely used, reused and redistributed by anyone, at any time and anywhere." The Charter has arisen from a conversation between governments and civil society, which has resulted in the promotion of the adoption of the six principles described below. Moreover, Open Data has been defined by the Open Knowledge Foundation as that which can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share-alike. Open Data core technical principles can be understood as follows.

• **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.

• **Reuse and Redistribution:** the data must be provided under terms that permit reuse and redistribution, including the 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.

The six principles of open data developed by the Open Data Charter are a globally-agreed set of aspirational norms for how to publish data, which can be summarised as follows.

1. Open by Default: This represents a real shift in how the government operates and how it interacts with citizens. At the moment, we often have to ask officials for the specific information we want. Open by default turns this on its head under the perspective that there should be a presumption of publication for all. Governments need to justify data that is kept closed, for example, for security or data protection reasons. To make this work, citizens must also feel confident that open data will not compromise their right to privacy.

2. Timely and Comprehensive: Open data is only valuable if it is still relevant. Getting information published quickly and in a comprehensive way is central to its potential for success. As much as possible, governments should provide data in its original, unmodified form.

3. Accessible and Usable: Ensuring that data is machine-readable and facilitates its dissemination, with portals being one way of achieving this. It is also important to consider the user experience of those accessing data, including such matters as the file formats in which information is provided. Data should be free of charge and under an open licence, as demonstrated by Creative Commons.

4. Comparable and Interoperable: Data has a multiplier effect. The more quality datasets you have access to, and the easier it is for them to talk to each other and hence, the more the potential value that can be acquired from them. Commonly agreed data standards play a crucial role in making this happen.

5. For Improved Governance & Citizen Engagement: Open data has the capacity to let citizens (and others in government) have a better idea of what officials and politicians are doing. Transparency can improve public services and help hold governments to account.

6. For Inclusive Development and Innovation: Finally, open data can help spur on inclusive economic development. For example, greater access to data can make farming more efficient, or it can be used to tackle climate change. Finally, we often think of open data as just about improving government performance, but there is a whole universe out there of entrepreneurs making money off the back of open data.

The Open Data Charter suggests a roadmap to implementation but what is more relevant to you is the benefits that they point out will imply the adoption of the principles. You can read more here. This would be of use when thinking of exploring open data and social innovation more closely. We will address open data for innovation further in this module, in unit 3.

The Government of Canada summarises the Benefits of Open Data as follows.

Support for innovation: Access to knowledge resources in the form of data supports innovation in the private sector by reducing duplication and promoting the reuse of existing resources.

Advancing the government's accountability and democratic reform: Increased access to government data and information provides the public with greater insight into government activities, service delivery, and use of tax dollars.

Leveraging public sector information to develop consumer and commercial products: Open and unrestricted access to scientific data for public interest purposes, particularly statistical, scientific, geographical, and environmental information, maximises its use and value, whilst the reuse of existing data in commercial applications improves time-to-market for businesses.

Better use of existing investment in broadband and community information infrastructure: Canada has invested in information and communications networks in the form of technical infrastructure and community services, such as libraries and social service agencies.

Support for research: Access to federal research data supports evidence-based primary research in the Canadian and international academic, public sector, and industry-based research communities. Access to collections of data, reports, publications, and artifacts held in federal institutions allows for the use of these collections by researchers.

Support informed decisions for consumers: Providing access to public sector service information to support informed decision-making, for example, real-time air travel statistics, can help travellers to choose an airline and understand the factors that can lead to flight delays.

Proactive Disclosure: Proactively providing data that is relevant to Canadians reduces the amount of access to information requests, email campaigns and media inquiries. This greatly reduces the administrative cost and burden associated with responding to such inquiries.

1 • 1.2 • Opening up data •

A dataset is a collection of organised data records where each element has the same structure, ordered for processing by a computer. For example, a dataset can be the list of schools in a country, the list of all state contracts for all its institutions, or the general budget of the nation.

The same dataset can have multiple distributions (or resources) that can vary in two dimensions as follows.

Temporal: in this case, the same data set has records associated with a data time. For example, the general budget of the nation has a different version each year, so too the list of contracts of a government.

Format: each data set can be represented in various formats. For example, if we consider that the list of government contracts can be represented in a table, it can be digitised to be opened with Acrobat Reader (in .pdf format), or Microsoft Excel (.xls), by any processor text (.csv) or processed by automated systems (.json), among other arrangements.

A wide variety of formats can be used to make data available to the public; however, not all meet the necessary requirements to define such data as "open". The format in which the information is published, that is, the digital base with which the information is stored, can, in fact, be open or closed. An open format is one in which the specificities of the software are available to anyone, free of charge, so that anyone can use them in the software itself without any limitation of reuse imposed through intellectual property rights. When, instead, the format is closed, it may mean that the format is proprietary and that the technical characteristics are not publicly available or that the file format is proprietary and, although the technical specifications are public, its use is limited.

The fundamental reason why it is important to clarify the meaning of "open" and why exactly use this definition can be summarised with one term: **interoperability**. This is the ability of different systems and organisations to work together. In our case, it is the ability to combine a database with others. Interoperability is the key that allows for the first practical advantage of openness: it increases exponentially the possibility of combining different databases and thus, developing new and better products and services.

Furthermore, the advantage of files in open formats is that this allows developers to produce software and services using these formats. This minimises the obstacles to reusing the information they contain. Using proprietary formats can lead to dependency on third-party software or the licensees of the formats. At worst, this

may mean that the information can be read using only a specific software format, which could be prohibitively expensive or become out of date over time.

Publishing data in open data portals in an efficient manner is key to developing strategies that address the following:

What data will be published iteratively and when? This refers to the roadmap to publishing the information. Given that there are generally limited resources, it is difficult initially to publish 100% of all the information available. So, it is important to have a roadmap in order to have clear and prioritised objectives in relation to what will be published. and when this will be achieved.

Where will the data be published? How will the data be published? This refers to the decision of the web address (the URL) where the open data portal will be, as well as the decisions regarding the formats in which the data will be published (JSON, CSV, JSON-LD). Some things that are important to consider are for example, if it will include an API for developers, or if massive downloads are expected.

What is the data update frequency? It should be acknowledged that there are datasets that need a higher update frequency than others, with some, for example, requiring daily (night, noon, etc.), weekly, monthly etc. updates.

Who is responsible for the publication of the data? This refers to those who are responsible for data management (system, institution, etc.). In all cases, it must be specified who publishes the data, and who is responsible for maintaining its accuracy and quality.

Who to contact if you have questions? It is important to explain clearly how to make inquiries relating to the data, in order to avoid misunderstandings.

What licence will be used to publish the data? The licence defines the permissions that the data owner grants in relation to what users can do. An open licence (for

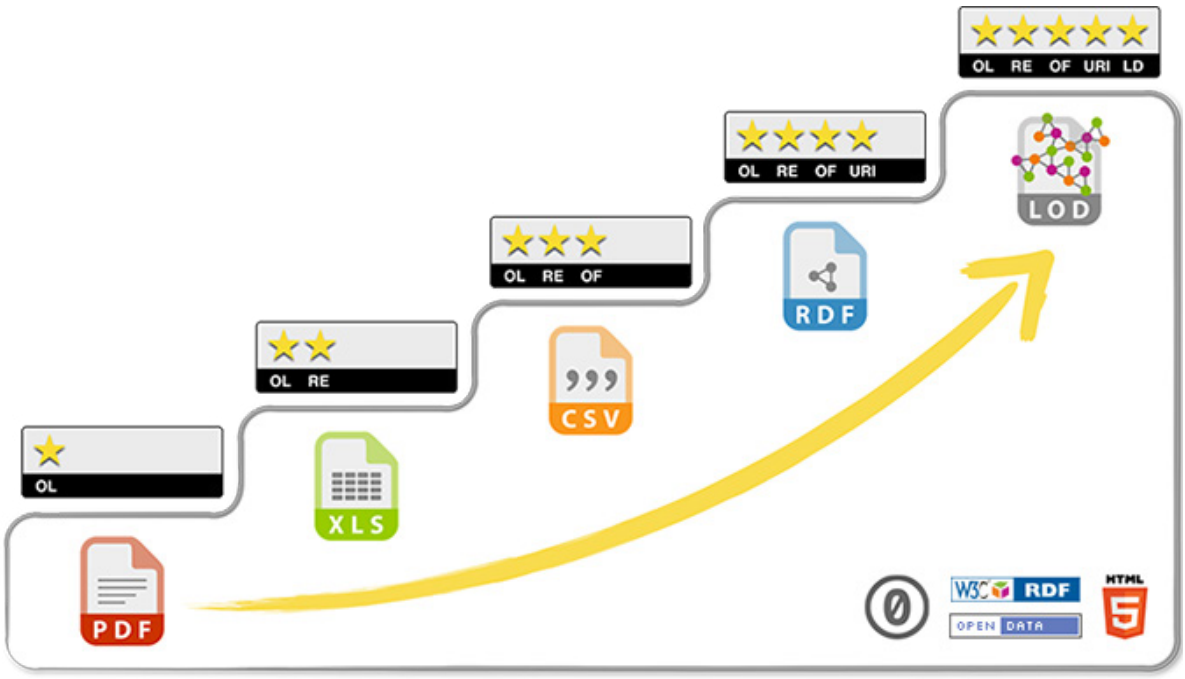
open data) must at least require attribution to the source and redistribution with the same licence.

Where can I find more reference information? There must be a place within the open data portal where you can access more information on related topics, such as where you can find data dictionaries, data manuals, or providing links to sites where you can find this data.

What is the regulatory framework of reference? It is very important always to have all the necessary references relating to policies, laws, decrees, resolutions, circulars, etc., which serve as a reference to everything that is exposed in the open data portal.

1 • 1.3 • Publishing open data •

The technical approach to data opening is based on the five-star data opening scheme defined by Tim Berners-Lee, a summary of which can be seen in the five star figure. This scheme proposes an incremental scale of data openness levels, where each level implies progress in terms of the objectives of open data: freedom of use, reuse and redistribution.



Five-star data opening scheme: Tim Berners-Lee (2012)

The great leap to the third star: the third star implies that the data is in a non-proprietary format, that is, it can be consumed and reused by anyone. To this end, the open data organisations are championing the standardisation of the open formats to be used in order to facilitate the work of data consumers. These formats are summarised in the following table.

Format	Description	Proposed Standard Scheme
CSV	Tabular data format where columns are delimited with a comma, although other separators such as semicolons are commonly accepted. Whilst it is not yet standardised, there are efforts to define good practices, such as RFC 4810.	JSON Table Schema
JSON	Data exchange format based on the key-value schema inspired by the Javascript object model. The main difference from the CSV format is the ability to define nested structures.	JSON Schema

Something to keep in mind is that depending on the types of data to be published, there are different formats to be used. For example, if the data is tabular, that is, it is contained in a table, one of the most used formats is CSV. On the other hand, if the data indicates geo-referencing there are other specialised formats to represent this information. Below are some of the most commonly used data types and formats.

Kind of data	Description	Common formats
Generic Data	Data that does not have specialised applications, and that normally corresponds to data from databases and reports, such as spreadsheets and information tables.	The characteristics of the formats to use are: 1. The use of standard, open (non-proprietary) formats should be considered. In certain justified cases (historical data, expensive transformations) proprietary formats can be used. 2. For traditional database and spreadsheet data, consider using CSV and JSON formats.
Images	Data that comes essentially from photographs.	1. You should use image file types based on the JPEG method and PNG files. 2. For vector and web display formats that require rendering, SVG (scalable vector graphics) is recommended.
Statistical data	Data normally used by specialised users of the statistical area. They are in particular formats for reuse and exploitation of specialised statistical applications.	As a data format, the Statistical Data and Metadata Exchange (SDMX) (OECD standard) is recommended. Acceptable for its widespread use are SPSS (PSPP) or STATA. In the latter case, you should also consider publishing a distribution of the data in CSV format.
Georeferenced data	Geographic data designed to capture, store, manipulate, analyse and display geographically referenced information in all its forms.	Among the most common formats we can find are: .kml, .geojson and .topojson.

When designing an open data policy, it is recommended to focus on the user, consult the demand for data and based on this prioritise the data to be published. When developing an open data opening plan, it is necessary that those who publish the data, be it the academy or the public sector, analyse and understand which are the sets that we can consider of high value or of greater relevance, in order to prioritise their publication according to certain factors, such as their value for user communities or to promote public participation. Other elements that may arise in cases of national or international contingency should also be considered, such as data on emergencies or natural disasters, epidemics or cases of corruption, which need to be published quickly.

1 • 2 • Introduction to open science •

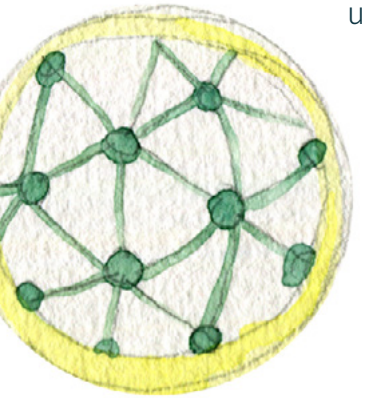
Open Science is just science done right

Jon Tennant (2018)

• Introduction •

Open science is the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of society, amateur or professional. Open science pertains to transparent and accessible knowledge that is shared and developed through collaborative networks. It encompasses practices, such as publishing open research, campaigning for ___ , encouraging scientists to practise open-notebook science, and generally making it easier to publish and communicate scientific knowledge.

According to FOSTER, Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. For the EU, Open Science, it represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. The idea captures a systemic change to the way science and research have been carried out for the last fifty years: shifting from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process.



The OECD defines Open Science as one that makes the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction. For UNESCO, the idea behind Open Science is to allow scientific information, data and outputs to be more widely accessible (Open Access) and more reliably harnessed (Open Data) with the active engagement of all the stakeholders (Open to Society).

By encouraging science to be more connected to societal needs and by promoting

equal opportunities for all (scientists, policy-makers and citizens), Open Science can be a true game changer in bridging the science, technology and innovation gaps between and within countries and fulfilling the human right to science.

The latest COVID-19 pandemic has shown how the availability of scientific information and research has made possible the acceleration of the vaccine against COVID-19 virus. This serves as clear evidence of the value of cooperation in the response to the Pandemic and at the same time it shows the dangers of treating evidence-based knowledge as an exclusive asset, or simple matter of opinion (United Nations, 2020). UNESCO is taking the lead in building a global consensus on values and principles for Open Science and contains reference to the first draft of the UNESCO (2020) Recommendation on Open Science that declares six aims and objectives including:

> 1. Universal access to scientific knowledge [as]... an essential prerequisite for human development and progress towards planetary sustainability.
>
> 2. Open Science sets a new paradigm for the scientific enterprise based on transparency, sharing and collaboration
>
> 3. As Open Science turns into a global movement, robust institutional and national Open Science policies and legal frameworks need to be developed by all nations to ensure that scientific knowledge, data and expertise are universally and openly accessible and their benefits universally and equitably shared.
>
> 4. This Recommendation outlines a common definition, shared values, principles and standards for Open Science at the international level and proposes a set of actions conducive to a fair and equitable Open Science transition at individual, institutional, national, regional and international levels.

In the Open Science MOOC, Open research data refers to the publishing of the data underpinning scientific research results so that they have no restrictions on their access. Openly sharing data opens it up to inspection and re-use, forms the basis for research verification and reproducibility, and opens up a path to broader collaboration.

1 • 2.1 • Open science principles •

In this section, materials from the Open Science Training Handbook and the Open Science MOOC are presented.

According to FOSTER, Open Science is about increased transparency, re-use, participation, cooperation, accountability and reproducibility for research. It is aimed at improving the quality and reliability of research through the principles of inclusion, fairness, equity, and sharing. Open Science can be viewed as research simply done properly, extending across the life and physical sciences, engineering, mathematics, social sciences, and humanities (Open Science MOOC). In practice, Open Science includes changes to the way science is done - including opening access to research publications, data sharing, open notebooks, transparency in research evaluation, ensuring the reproducibility of research (where possible), transparency in research methods, open source coding, software and infrastructure, citizen science and open educational resources.

One of the key elements of Open Science is reproducibility, which pertains to research **data and codes being made available to others**, who are able to obtain the same results as ascertained in scientific outputs. Closely related is the concept of replicability; the act of repeating a scientific methodology to reach similar conclusions. These concepts are core elements of empirical research.

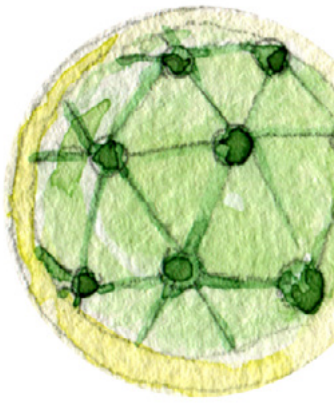
Improving reproducibility leads to increased rigour and quality of scientific outputs and thus, to greater trust in science. There has been a growing need and willingness to expose research workflows from initiation of a project and data collection right through to the interpretation and reporting of results. These developments have come with their own sets of challenges, including designing integrated research workflows that can be adopted by collaborators, while maintaining high standards of integrity.

The concept of reproducibility means being able to do or apply a method again in another piece of research. It is directly applied to the scientific method by providing clear and open documentation, thus making the study transparent and reproducible.

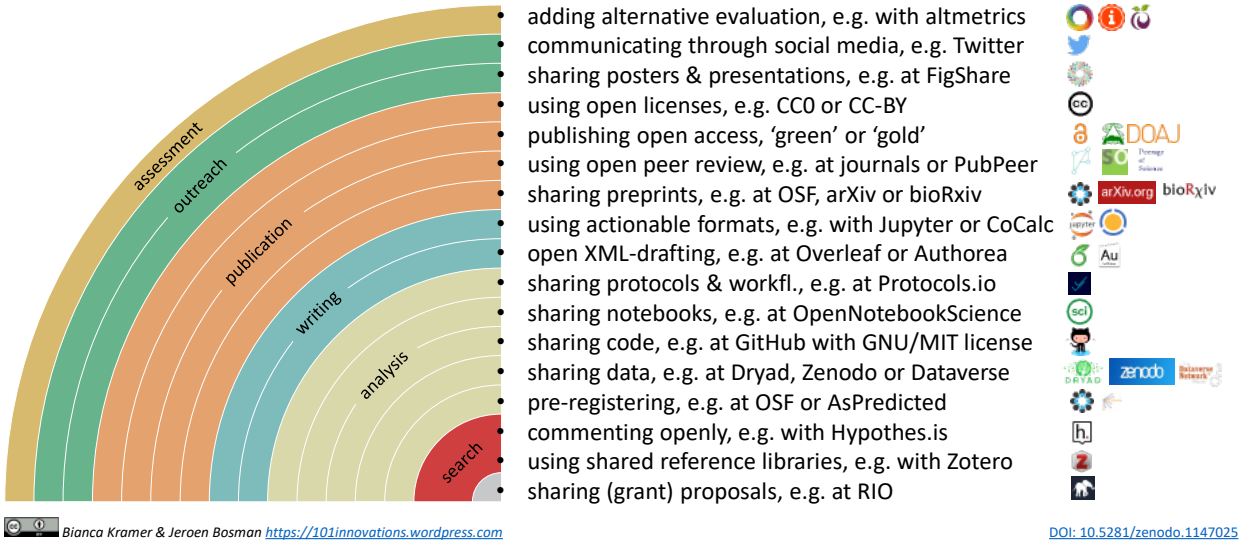
Goodman, Fanelli, & Ioannidis (2016) note that, in epidemiology, computational biology, economics, and clinical trials, reproducibility is often defined as the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

This is distinct from replicability, which refers to the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected. A simpler way of thinking about this might be that reproducibility is methods-oriented, whereas replicability is results-oriented.

Reproducibility can be assessed at several different levels: an individual project (e.g. a paper, an experiment, a method or a dataset), an individual researcher, a lab or research group, an institution, or even a research field. Slightly different kinds of criteria and points of assessment might apply to these different levels. For example, an institution upholds reproducibility practices, if it institutes policies that reward researchers who conduct reproducible research. Further, a research field might be considered to have a higher level of reproducibility, if it develops community-maintained resources that promote and enable reproducible research practices, such as data repositories, or common data-sharing standards.



Rainbow of open science practices: You can make your workflow more open by...



It is key to critically enabling open science practices in teaching and learning because it fosters research-driven learning opportunities through open data, which is a condition sine qua non for reproducibility and scientific progress, facilitating reuse. For Ioannidis and Khoury (2011), Opening up data enables to detect false claims and inaccuracies and allows for replicability tests. In sum, opening up research data can have a wide societal impact.

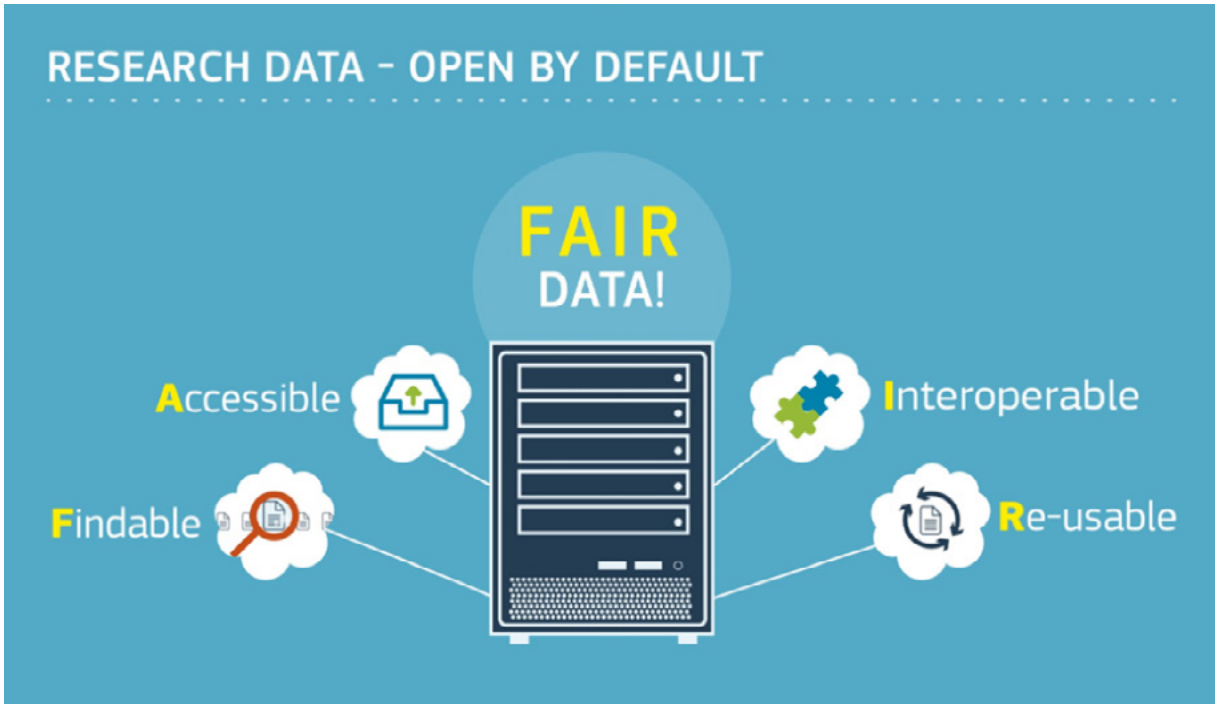
1 • 2.2 • FAIR data principles •

According to LIBER EU, The FAIR Data Principles is a set of guiding principles aimed at making data findable, accessible, interoperable and reusable. It provides guidance for scientific data management and stewardship that is relevant to all stakeholders in the current digital ecosystem. They directly encourage data producers and data publishers to promote maximum use of research data.

During the Lorentz Workshop Jointly Designing a Data FAIRport (2014), participants formulated the FAIR data vision to optimise data sharing and reuse by humans and machines, which resulted in the publication of The FAIR guiding principles for scientific data management and stewardship in "Scientific Data". The FAIR

principles can be understood as **Findability**, **Accessibility**, **Interoperability**, and **Reuse** of digital assets. The principles emphasise machine-actionability (i.e. the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention), because humans are increasingly relying on computational support to deal with data as a result of the increase in volume, complexity and speed of creation.

For OpenAire, the FAIR principles are needed to increase the availability of online resources, thus, data need to be created with longevity in mind. Providing other researchers with access to your data facilitates knowledge discovery and improves research transparency. The FAIR principles describe how research outputs should be organised so they can be more easily accessed, understood, exchanged and reused. Major funding bodies, including the European Commission, promote FAIR data to maximise the integrity and impact of their research investment.



https://www.openaire.eu/how-to-make-your-data-fair / creativecommons licenses

The EU proposes six Recommendations for Implementation of FAIR practice, which can be summarised as follows.

1. Fund awareness-raising, training, education and community-specific support.

2. Fund development, adoption and maintenance of community standards, tools and infrastructure.

3. Incentivise development of community governance.

4. Translate FAIR guidelines for other digital objects.

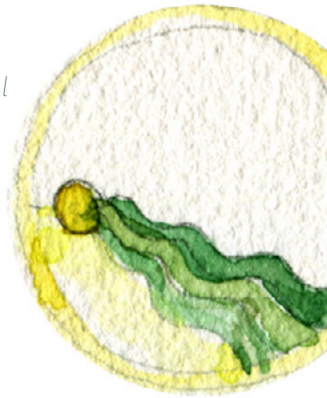
5. Reward and recognise improvements in FAIR practice.

6. Develop and monitor adequate policies for FAIR data and research objects.

It is important to promote open science ethical principles and practices in research education to promote scientific data creation and sharing. This can be done by mapping and showcasing the impact open science can have to encourage transparent and democratic research practices.

1 • 3 • Open data and innovation •

Contributor: Carla Bonina



In this model of social change, innovators co-create solutions for a wide variety of social problems. These solutions are openly available to other innovators and users, who can then act to further improve and advance them.

Bonina, López-Berzosa, Scarlata 2020

• Introduction •

The value of open data is realised through its use. However, capturing attention and driving innovation across the spectrum of data users, intermediaries and consumers is challenging. How to encourage participation in the demand for open data? How to generate and grow the community of data users? This module is aimed at explaining what kind of value can be created with open data, who are the main groups of users of the data and how to plan strategies to encourage both the participation of those groups and the generation of new initiatives with open data.

1 • 3.1 • The value of open data •

The value of open data lies in its use. This first unit aims to cover the following aspects:

- What kinds of benefits can open data and civic technology generate?
- What examples can be found in the world?
- How are these benefits generated?
- What is the role of participation and collaboration for the generation of value?

In a world with increasing poverty, inequality, environmental degradation and injustice, digital innovation with a social focus appears as a ray of hope to solve such problems. **Civic technology** is a type of social digital innovation, in which digital technologies are implemented to improve the relationship between citizens and the government in order to include more people in public decision-making, citizen empowerment, or to improve access and delivery of public services. In many cases, in particular at a regional level, open data is an input for the development of innovations based on civic technology.

Open data opens up new possibilities to generate value, both in the economic, social, environmental, and democratic fields. There are several open data projects that have a civic or social objective that at the same time may be generating economic benefits. While the types of benefits that open data can bring are not necessarily exclusive, we present them separately below for simplicity.

Economic Benefits: On the economic side, as companies, governments and citizens use or reuse open data, new products, processes and business opportunities emerge. In 2011, a study commissioned by the European Commission estimated that the economic value of opening up and reusing public sector information was approximately € 40 million per year for the European Union alone. On a global scale, McKinsey estimated in 2013 that open data could contribute to the generation of between $ 3 and $ 5 trillion per year in the global economy. Whilst these figures are still speculative, they open up a range of possibilities when it comes to generating economic value.

Social and democratic benefits of civic technology and open data: Equally important, open data is seen as particularly beneficial in increasing transparency, fighting corruption, and promoting social inclusion, but the value of these benefits is much more difficult to assess in monetary terms. This perspective of value has been particularly important in Latin America, where the lack of transparency and accountability, as well as social needs, have been at the centre of the agenda of the open data movement in the region. Concern about transparency and accountability is not, however, a problem

exclusively for developing countries. Open data policies in pioneer countries, for example, the UK, were partly the result of a lack of trust in the political system and its accountability.

There are also many cases in which open data is built on the basis of volunteer work, cooperation and civic engagement. There are open data initiatives of collaborative peer production based on the public domain, in which participants collaborate, contribute their time and exchange knowledge and experience to achieve a social or public interest good. A pioneering initiative in the field of geographic location is the Open Street Map, a collaborative project to create a free editable world map. Similar to Wikipedia, Open Street Map is produced by a community of local knowledgeable contributors, who update and maintain data on routes, train stations, cafes and much more around the world, all in an open format. This project has inherent utilitarian value - anyone with a GPS can download and use it for free, and it has proven particularly useful in remote areas. However, this common task has a higher value. As in other cases of open data, Open Street Map can be used as a platform to create other services based on mapping data, such as transport, humanitarian aid or monitoring.

Participation, collaboration and co-creation with civic technology and open data: Both, civic technology and open data, promise a variety of benefits economically as well as socially, politically and democratically. The potential of open data not only at a regional level but also globally, is particularly facilitated by greater access and affordability to the Internet, computers and mobile telephony, among other digital technologies. The generation and appropriation of its benefits result from the participation, collaboration or co-creation of services by a diversity of actors, two fundamental pillars in the framework of Open Government.

The concepts of participation, collaboration and co-creation are undoubtedly related. For, public or civic participation can lead to cases of collaboration or co-creation of services. For example, an open session between the government, civil society organisations and citizens can give ideas to prioritise a series of data to be opened up, which, over time, can also lead to concrete initiatives between a civil

society organisation and the government. In other cases, an open session can be a good input to modernise a government service, which is not necessarily co-created in collaboration with an outsider.

In simple terms, we understand co-creation as pertaining to those processes or activities where at least two actors (e.g, public, private, governmental or civic) collaborate in the realisation of a project to achieve a certain result. In other words, it refers to where more than one part is required for value creation.

1 • 3.2 • Open data for innovation •

Despite expectations, realising the potential benefits of open data is complex. Part of the problem governments face is the need to cultivate an ecosystem of reusers. On the demand side, the authority responsible for the open data platform needs to cultivate a base of independent reusers, either to contribute new services or applications based on open data, as well as end users who consume these services. A key challenge that authorities of open data platforms face is governance, that is, how to attract and keep active those who have the capacity to generate new services and applications for the economy. This unit introduces a simple model to foster innovation with open data, based on rules and tools.

The logic of the contribution of open data to the generation of economic benefits can be summarised as pertaining to two types as follows.

• Direct benefits: include the generation of income from new products or services, the creation of jobs and/or the reduction of economic costs. An example would be a new company that earns its revenue from visualising or transforming open data for the financial industry.

• Indirect benefits: the reduction of waiting times for users of applications with open data, the increase in search options for decision making, the increase in the

efficiency of the public sector and the growth of related industries, among others. An example of indirect benefits would be the reduction of hours lost in traffic by the inhabitants of a region, based on the use of open data regarding transport.

To cultivate an ecosystem of entrepreneurs and innovators around open data, the innovation platform approach is useful. It is key to having tools for fostering innovation with open data and working with a series of priorities or public problems to be solved. In order to cultivate the demand of entrepreneurs and innovators with open data, among the various social tools that can be used, it is possible to distinguish two types: information and software, and social.

1. Information tools and software: These are the most obvious resources on an open data platform, and comprise data sets as well as other information tools, such as manuals and related videos for use. They also include software tools, such as APIs and web portals, that allow developers and innovators to access data sets programmatically.

2. Social tools: These tools complement those of information and software, being those that involve the participation of demand, through the deployment of events, financing or other types of incentives, of which three stand out.

• **Hackathons:** These are marathons of ideas, design and software prototyping, bringing together specialists, technical programmers, designers and entrepreneurs to work collaboratively. The sessions usually last between 24 and 48 hours, having specific challenges to guide the event, and in many cases they also have parallel workshops for capacity building. Hackathons are often one of the first strategies to promote the reuse of open data — and at the same time they create or strengthen the community, especially of entrepreneur

• **App challenges:** These competitions are similar to hackathons, but typically highlight the competitive element with prizes or recognition awarded by a panel of experts or judges, and can last longer than a hackathon. There are competition rules, which specify the products or applications considered “eligible” for the awards.

• **Incubation processes and funds:** These tools are used to incubate well-formulated project that have potential for the development of an application, product or service based on open data. For this purpose, seed financing is offered to build and apply scalable models as these programmes may be more successful in generating impact with open data. In reality, there are very few examples in Latin America regarding such programmes.

To ensure that Open Data properly benefits society, it is key to promote more channels of collaboration between citizens, business and government. Opening up data without a community using it will not generate value per se. It is common to find that public officials report difficulties in prioritising those data that may be of greatest interest to the groups that make up the demand for data. Facilitating spaces for collaboration between the government and these actors is a strategy that has proven to be effective when creating and promoting the use of open data. The third wave of data is about data access and reuse. Here you can find a wealth of information and toolkits to explore issues and advice concerning the third wave of data. Also, it is important to invest in understanding open data business models. Whilst technology entrepreneurs, startups and private sector companies are very relevant actors in the open data ecosystem, they are still disconnected or even absent in open data discussions. This is because it is not easy to generate business with open resources hence, emphasis needs to be placed on understanding what kinds of business models are viable or sustainable in the ecosystem.

Back to Understanding Data: Praxis + Politics

Illustrations, editorial and graphic design