

Open Government Data

Reflections on the course about data management and open government data at the university of applied science Bern, held on August 24-26, 2023

Damian Oswald

September 11, 2023

This report reflects on the concept of open government data. The article goes on to explain principles behind OGD such as the FAIR principle, the five-star system, open standards, interoperability, and roles of data owners, custodians, and stewards. Subsequently, the Swiss OGD ecosystem is highlighted, exploring the legal framework and regulations surrounding it.

“The world’s most valuable resource is no longer oil, but data” – at least that’s how *The Economist* phrased it in an article in 2017 (1). But is that true? There are certainly good reasons to support this metaphor. Much like oil, data can also be seen as a raw material – and as a raw material they are worthless unless processed.

However, there is the argument that data could be viewed more like an infrastructure than oil in the sense that it provides the fundamental framework or system needed for things to function. This could be in the context of operations, processes or even economies. Data, like infrastructure, is essential for information flow, decision making, and overall functioning in today’s digital era. And, unlike oil, which is consumed and needs to be replenished, data can be used repeatedly. Seeing data as a sort of infrastructure is used as an argument that its publication should be a responsibility of the state — in the form of open government data (2).

Open government data (OGD) refers to the idea that data collected and generated by government entities should be freely accessible to the public, excluding information that compromises privacy and security. This type of data is often used as a means to promote transparency and accountability in government, enable public participation, and drive economic growth through the development of new apps, products, and services. It also offers potential for innovation by providing raw material for research, as well as the basis for planning in areas like transport, healthcare, or environmental policy.

Some principles behind OGD

The FAIR principle

OGD should be FAIR. The latter stands for **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse. *Findability* means that data have a unique and persistent identification. For example, personal data should be given identification numbers (ID) that can be uniquely assigned to individuals. Furthermore, datasets should be described with *metadata*. *Accessibility* refers to the principle that

data are freely available via a URL. If data is protected, then at least their metadata should be published. *Interoperability* refers to the publication of data and metadata in a known (and widespread) format. Lastly, by *Reuse*, we mean that the provenance of data is indicated, the data is always provided with data usage license and is retrievable in a sensible size.

The five-star system

Not all open data are of the same quality – *all OGD are open, but some are more open than others*. In 2010, Sir Tim Berners-Lee suggested a 5-star ranking for open data. The data get more stars as proprietary formats are removed and links are added (table 1).

Open standards

For data to be truly open, it is not enough that it is accessible. The format also needs to follow an *open standard*, that is, a standard that is widely used, internationally accepted, consensus-based, in the public domain and clearly defined. Importantly, metadata, i.e., data about the data, should also follow that open standard. Examples of that are *Dublin Core*, *Schema.org*, or *DCAT*.

Interoperability

Interoperability is the ability of independent, heterogeneous computer systems to work together as seamlessly as possible in order to exchange information in an efficient and usable manner, and to make it available to the user without the need for special agreements between the systems. For this, compliance with common standards is necessary.

Data owner, custodian and steward

The roles of data owners, data custodians, and data stewards, all play critical parts in the management and governance of data, but their responsibilities are distinct:

1. **Data Owner:** This is usually a member of management who has legal authority and control over some

Stars	Idea	Example format
★	Make your stuff available on the web under an open license (no matter the format)	PDF
★★	Make it available as structured data (e.g., Excel instead of image scan of a table)	XLS
★★★	Make it available in a non-proprietary open format (e.g., CSV instead of Excel)	CSV, JSON, XML
★★★★	Use URIs to denote things, so that people can point at your stuff	RDF
★★★★★	Link your data to other data to provide context	LOD

Table 1: The five-star system according to Tim Berners-Lee. The more stars, the more *open* the published data is.

data. They are responsible for the security and quality of the data and for determining who should have access to the information. They must also ensure compliance with relevant legal and regulatory requirements.

2. **Data Custodian:** The data custodian is typically responsible for the technical environment and operations needed to protect and maintain data. These responsibilities may include data storage, retrieval, archival, or purging. The data custodian ensures data is managed and protected, execute the access decisions made by the data owner, and maintain the data's technical infrastructure.
3. **Data Steward:** The data steward ensures the quality and usability of the data. They concern themselves with data standards, data definitions, data quality metrics, and business metadata. They serve as a bridge between data owners (who decide the policies) and data custodians (who implement the policies) by making sure that data is reliable, understood, and used correctly (3).

The Swiss OGD ecosystem

Switzerland passed the Freedom of Information Act and Regulation (*Öffentlichkeitsgesetz, Öffentlichkeitsverordnung*) on July 1, 2006. These laws state, that anybody has the right to get access to non-protected governmental documents (4). Although that was a first step towards a more transparent government, it's not open government data. The [parliamentary group for digital sustainability](#) (Parldigi), founded in 2009, is concerned with promoting concepts such as open source software, open standards, open data, open access, open content, open internet, open government in parliament. Outside of parliament, the association [opendata.ch](#), founded in 2011, has been committed to strengthening transparency, participation and innovation in Switzerland with respect to data.

The legal framework around OGD recently experienced renewal with the OGD strategy Switzerland 2019-2023 (2). This strategy creates a business office OGD within the federal statistical office responsible with its implementation. Goal is the targeted publication of government data at a high quality. For this purpose, the meta

data platform [opendata.swiss](#) is created. Supporting that strategy, the federal law on the use of electronic means for the fulfillment of official duties (EMBAG), which is expected to come into effect in the beginning of 2024, will force all governmental software and data to be published *by default* (5). Specifically, the law states:

Art. 10 Open Government Data

¹ *Die diesem Gesetz unterstehenden Verwaltungseinheiten machen ihre Daten, die sie zur Erfüllung ihrer gesetzlichen Aufgaben beschaffen oder generieren und die elektronisch gespeichert und in Sammlungen strukturiert vorliegen, öffentlich zugänglich.*

However, there are many exceptions to that law, such that finding excuses not to publish data for governmental bodies will remain relatively easy. In addition, the data protection act states that for the publication of *personal data*, i.e., data referring to any legal entity, a special legal basis is required – thus, personal data are specifically exempt from the EMBAG (6).

How to anonymize personal data

Personal data may only be published if it is anonymized. One such method for anonymization is *k*-anonymity. Its protocol was developed in 1998 by Samarati and Sweeney (7). The (basic) algorithm to determine the degree of anonymity *k* of some personal data is relatively simple. In R, it could be written as follows:

```
K <- numeric(nrow(D))
for (i in 1:nrow(D)) {
  for (j in 1:nrow(D)) {
    if(i==j) next
    if(all(D[i,qi]==D[j,qi])) K[j] <- K[j]+1
  }
}
k <- min(K)
```

Here, *K* is the number of identical entries found for an observation in the data set *D*, while *k* is the minimum of vector *K*. The vector *qi* contains variable names of the quasi-identifiers. Standard personal data should only be published when $k \geq 20$, and sensitive personal data only if $k \geq 50$.

The impact of open government data

Open government data can have all sorts of benefits. This includes (but is not limited to) the following:

1. **Transparency:** Open government data allows for greater transparency in public administration. It enables citizens to see what decisions are being made and how resources are being allocated. This may build trust, as citizens can see the rationale behind policies and decisions, and helps to hold government officials accountable.
2. **Innovation:** Open data can be used by individuals or organizations to create innovative solutions and services. As such, it can foster entrepreneurship and stimulate economic growth.
3. **Sustainability:** Open government data can be useful in academic research, and may be used as a resource in education, both for teaching and learning. Students can analyze real data for projects or assignments. It also encourages citizen participation and engagement in government processes by enabling more informed discussions and debates.
4. **Efficiency:** Open data fosters collaboration between different government entities, as well as between governments and citizens, nonprofits, businesses, etc. This can generally lead to a more efficient and attractive government administration.

This all sounds great. However, individual governmental bodies need to know what their specific goals are, i.e., why they are publishing data. And ideally, they should (quantitatively) measure the impact in that regard, for example by applying the social return on investment approach (8). Unfortunately, determining causal effects like this is very difficult. It might be much more straightforward to pose surrogate questions: What businesses rely on the open data? How many research papers were published using the data? How often was the data used by journalists? Did the transparency lead to any citizen participation? So really, we are interested in the intermediaries, the entities using the open data, rather than their final effects.

Conclusion

Open government data represents a significant opportunity to foster transparency, accountability, innovation and citizen participation in governance. It also bolsters economic growth and encourages data-driven policy making. In Switzerland, laws start to come into effect that specifically promote open government data. However, for OGD to be truly open (in spirit), some concepts such as FAIR, the five-star system, interoperability and open standards need to be followed. Also, it's crucial

for governments to clearly define their OGD goals and assess their impact, ensuring that the data is effectively serving its intended purpose and benefiting society.

References

1. The Economist, The world's most valuable resource is no longer oil, but data (2017).
2. Schweizerische Eidgenossenschaft, Strategie für offene Verwaltungsdaten in der Schweiz 2019-2023 (Open-Government-Data-Strategie, OGD-Strategie) (2018), (available at <https://fedlex.data.admin.ch/eli/fga/2019/125>).
3. D. Plotkin, *Data stewardship: An actionable guide to effective data management and data governance* (Academic press, 2020).
4. Schweizerische Eidgenossenschaft, Bundesgesetz über das Öffentlichkeitsprinzip der Verwaltung (Öffentlichkeitsgesetz, BGÖ) (2004), (available at <https://fedlex.data.admin.ch/eli/fga/2004/1275>).
5. Schweizerische Eidgenossenschaft, Bundesgesetz über den Einsatz elektronischer Mittel zur Erfüllung von Behördenaufgaben (EM BAG) (2023), (available at <https://fedlex.data.admin.ch/eli/fga/2023/787>).
6. Schweizerische Eidgenossenschaft, Bundesgesetz über den Datenschutz (Datenschutzgesetz, DSG) (2023), (available at <https://fedlex.data.admin.ch/eli/cc/2022/491>).
7. P. Samarati, L. Sweeney, Protecting privacy when disclosing information: K-anonymity and its enforcement through generalization and suppression (1998).
8. M. Stuermer, M. M. Dapp, in *2016 conference for e-democracy and open government (CeDEM)* (IEEE, 2016), pp. 197–203.