March 22, 2024

**Submission>
Pollution Information Portals and the Right to Know: Strengthening access
to information on releases of hazardous substances**

**Recommendations from the Open Knowledge Foundation**

Answering the public call for submissions and drawing on our two-decade experience, the Open Knowledge Foundation wants to contribute to the work of the Special Rapporteur by providing five recommendations that would make Pollution Information Data Portals interoperable, sustainable, accountable, resilient and adaptable to technological changes. At the Open Knowledge Foundation, we have proven how adopting free and open software, open data and open standards, with active communities on the ground, can guarantee fair access to information and enhance its benefits.

As part of the contribution to the work of the current call of submissions issued by the United Nations, we want to put forward recommendations that would contribute to strengthening access to information on releases of hazardous substances.

* **Recommendation One**: adopt openness as a design principle for all the Pollution Information Data Portals.

This means intentionally designing processes, policies, standards and protocols in a way that anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness). By adopting openness as a design principle for Pollution Information Data Portals, the authorisation for use and reuse of the information provided by the data portals will be given in advance. Furthermore, each layer of the Pollution Information Data Portal will be interoperable, reproducible and generative. By adopting openness at all levels, it would also make the effort more transparent, accountable and accessible.
* **Recommendation Two:** develop and adopt open standards for Pollution Information Data Portals.

By adopting domain-specific open standards for pollution information data portals, all the data will be generated, validated and shared better. One of the key issues to solve for Pollution Information Data Portals is data frictions. There is significant friction in working with data. This friction manifests itself in a variety of ways:

* *Access:* Data is often published in repositories or catalogues that have restrictive access, for example, behind commercial licenses with strict rules on reuse.
* *Discoverability:* Data is hard to find. This can be especially true in terms of the discovery of contextual data to support a main line of inquiry.
* *Structure:* Source data often has errors, is badly formatted, is inconsistent, and in general requires significant manual processing to make usable.
* *Standardisation:* Data is stored and archived in different formats, including platform-specific or difficult-to-use formats, poorly described and structured, thereby incomplete.
* *Semantics:* The “same” data from different sources often employs radically different semantics, resulting in issues around connecting data across domains.
* *Tooling:* There is very little commonality in data formats across tooling, especially when a lot of data work is multifaceted and uses a mixture of consumer tools like spreadsheets, and professional tools such as programming in Python or R. This results in large amounts of manual and semi-automated data transfer across tools.

A tangible help in building Pollution Information Data Portals could be [Frictionless Data](https://frictionlessdata.io/), an open-source toolkit made of standards for data and metadata interoperability and software tools for data management powered by those standards. Frictionless standards provide a simple way to package data and improve its portability. Thanks to the modular approach, they are also very easily extensible and customisable to accommodate domain-specific requirements. When it comes to data manipulation, the software implementations of those standards offer a perfect solution to aim at simplifying your data workflow, and ensuring data quality. Allowing data to be described, extracted, validated and transformed in a reproducible way, they make all data curation activities completely transparent and traceable, and improve the efficiency and consistency across data management, support many data sources and formats, and provide popular platform integrations.

* **Recommendation Three:** adopt free software to manage the data and invest in lightweight free software solutions for localised efforts.

Free Software can be used for any purpose and is free of restrictions such as licence expiry, geographic limitations, or unilateral sanctions. Anyone can study Free Software and its code without non‐disclosure agreements, trade secrets, or similar restrictions. Free Software can be shared and copied at virtually no cost. Anyone can modify Free Software, and these improvements can be shared publicly. All those advantages will reduce the financial and environmental costs while improving the localisation of the efforts into different languages and usability needs, connectivity, skills and systems adopted.

An example of free software used in data portals is CKAN, a Digital Public Good that is an open-source DMS (data management system) for powering data hubs and data portals. CKAN makes it easy to publish, share and use data. It powers [catalog.data.gov](http://catalog.data.gov), [open.canada.ca/data](http://open.canada.ca/data), [data.humdata.org](http://data.humdata.org) among many other sites and it is a good solution for a sophisticated pollution information portal.

* **Recommendation Four:** develop local digital skills in the teams and engage with the active open data community locally and globally.

For a successful implementation of the Pollution Information data portals, increasing the skills of the civil servants is a *sine qua non* condition and we would like to refer to the competencies framework for [Artificial intelligence and digital transformation (UNESCO, 2023)](https://unesdoc.unesco.org/ark%3A/48223/pf0000383325.locale%3Den) which offers a comprehensive path to develop the skills those in charge of the pollution information infrastructure need in the near and mid-term.

However, a robust and skilled public sector is not enough. For its purpose to be fulfilled, the supply of data provided by Pollution Information Portals needs to be actioned. Toxics information will only increase shortly, both in quantity and complexity, as well as dispersion. More authorities (whether at Federal, State, City or County levels) are making commitments to publish the data they collect, to increase transparency and tackle the environmental crisis together with citizens, more citizens are generating their own data sets and contrasting them with the information available by public bodies. We recommend this feedback loop between government and citizens be catalysed by empowering media, civil society and individual changemakers to understand and use data that has been released and by sharing the data that is generated by civil society efforts. And that requires a data-literate society. People need skills to find, clean, interpret and analyse data to find evidence, create visualisations and tell stories so that they can present their arguments with more force and improve public policies. While the efforts are still dispersed, a global movement around open data could help both with technical and policy aspects of Pollution Information Data Portals.

One example of engagement at a global level at the intersection of data and civil society is the [Open Data Day](https://opendataday.org/), a bottom-up initiative where institutions, groups and communities dedicate a day to reach out to new people and build new solutions using open data. The organizations involved are every year inviting the technical community to become active in tackling social issues, inviting the public sector to update their data literacy, and building collaborations with the private sector and academia. Such events are catalysers of lasting collaborations and improve the communication and exchange of experiences between those using public interest datasets and those publishing them.

**About Open Knowledge Foundation**

Open Knowledge Foundation is a global non-profit organisation focused on realising open data’s value to society by helping individuals and organisations access and use data to address social problems. The Open Knowledge Foundation team has several decades of combined experience working on the legal, technical, and social aspects of data sharing.

Open Knowledge Foundation is widely known to be one of the originators and leaders in open data and in building robust, sustainable open infrastructure for publishing, sharing and working with data generally.

The organisation and team have a combination of three key characteristics that will enable the project’s success:

* Open Knowledge Foundation has practical experience delivering world-class data software and tooling on time and on budget.
* Open Knowledge Foundation has extensive experience in developing and delivering data literacy training to a wide variety of communities.
* Open Knowledge Foundation has strong connections with key communities necessary for ensuring that the results of the project are taken up and adopted by data users as well as extensive experience working with a diverse set of stakeholders to deliver sustainable open source data infrastructure that meets the needs of its users.

Open Knowledge Foundation has successfully executed and sustained several large and small technical projects, all of which demonstrate our capacity to deliver key open source data infrastructure. Furthermore, the team has extensive prior experience working with specific data-sharing communities in research, government and the civic society space.

Notable projects include [​CKAN​](https://ckan.org/), the world-leading open source data management and publication platform, recently registered as a Digital Public Good, [Frictionless Data,](https://frictionlessdata.io/) a progressive open-source framework for building data infrastructure, and the ​[Global Open Data Index](https://index.okfn.org/)​ for tracking the state of government data.