Best Practices in Open Data for the Environment

James Smith, Cleanweb UK

Abstract

This paper presents an overview of Open Data, with a summary of licensing, best practices, and its application to the environmental data arena.

What is Open Data?

Open Data is a very simple concept; it is simply data that can be used without restriction by any user. The Open Knowledge Foundation's Open Definition summarises it as:

> *"A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike."*

Open Data is fundamentally about removal of restrictions to reuse, and is a legal definition as well as a technical one. The Open Definition includes a comprehensive guide to data licensing for more details.

It is important not to confuse Open Data with other popular terms, such as Big Data or Linked Data, or with the things that are built *using* Open Data, such as web or mobile applications. These are all additions to, or layers on top of, Open Data.

Complementary to Open Data are efforts like Open Source software, which enables the free sharing of source code (including models and so on), or Open Research, which covers the methodologies around collection and processing of data.

What sort of data do you mean?

Open Data can cover any type of dataset, from transport timetables to scientific measurements. However, particular sets of interest in the environmental space include:

- Scientific climate data & temperature records
- Weather & forecasts
- Natural disasters
- Energy subsidies
- Energy potentials
- Realtime energy production and pricing data

- Standardised regular reporting of national & local emissions data
- Biodiversity, including deforestation & marine stocks
- Food security & production

Why should we make our data Open?

Innovation

If datasets can be easily accessed and legally reused, then innovation can flourish on top of them. Deloitte have [stated](#) that:

> *"open data … will be a vital driver for growth, ingenuity and innovation in the … economy"*

This innovation can take many forms, both commercial and non-commercial:

- Data cleaning
- Transformation to alternative formats
- Crowdsourcing of extra data or metadata
- Linking and tagging of disparate datasets (using Linked Data approaches)
- Large scale analysis (using Big Data techniques)
- Applications, from simple data display to deeper cross-set analysis
- Data journalism and visualisation
- Discovery of new efficiencies from combined datasets

A single organisation may build applications on top of its data for a particular purpose, but if that data is Open, then many organisations are able to build different applications for different users. As no single tool will fit the needs of all users, taking an Open Data approach allows much wider use of the data to fit with wider user requirements.

Accountability & Transparency

Opening data has many other advantages beyond innovation:

- Greatly improved transparency and accountability
- A more informed citizenry
- Increased trust in organisations opening their data, and in their practices
- Better measures of social and environmental ROI
- Improved economic efficiency
- Reduced corruption

At the international level, the [World Bank](#) has recognised the power of Open Data to unlock knowledge and information, raise levels of development, and reduce poverty worldwide:

> *"Open access to data is a key part of the World Bank's commitment to sharing our knowledge to improve people's lives."*

The [Open Government Partnership](#) is also working to promote transparency in governments across the world, with major Open Data efforts starting up in many countries as an essential component in increased accountability and elimination of corruption.

Environmental Applications

The environmental sector is in a unique position where scientific conclusions require societal change. Mistrust and disinformation are rife, and are delaying substantive action on the required timescales.

By taking an Open Data approach, organisations and institutions can:
- become more transparent and accountable, reducing mistrust in the data, the conclusions, and in themselves.
- stimulate innovation, which is essential for rapid societal or technological change.
- improve public communication, by allowing enabling a wide range of approaches to communicating scientific data and conclusions.
- help politicians to build the will to make required changes, by giving them reliable and transparent data to back up their arguments.

How do we open up data?

In order to make a piece of data open, only the following steps are necessary:

1. Collect the data together into a single file or set of files.
2. Resolve any licensing conflicts; for instance, different parts of the same dataset may be owned by different parties.
3. Choose an open data license, such as the [Open Data Commons Attribution License](#). Alternatives, including share-alike licenses are available from the [Open Data Commons](#) site.
4. Upload the data to a publicly-accessible part of your website. Registration should not be required to access the data.
5. Include the following information along with the data:
   a. License details (from step 3)
   b. Technical details of the format that the data is stored in. Note that this does not *need* to be a standard format, as long as you can explain it to users, but it should *not* be in formats that require proprietary software to use (such as XLS), or in non-machine-readable formats (such as PDF).
   c. Details of when the data was last updated, and will be updated next.
   d. Provenance of the data, including details of original creator.
   e. Methodology information, such as how data was collected, calibrated, and transformed prior to upload.

Note that there is no requirement above to transform the data yourselves, or to convert it into a particular format (unless it is currently in a format that requires proprietary software). However, *good* open data will be in a standard structured format such as XML, CSV or JSON. See the technical openness section of the Open Data Handbook for more details.

You may also wish to consider layering other services around the data, but these are certainly not *required*.
- Storing in or linking from a third-party data *hub* (like thedatahub.org), to allow users to find it more easily.
- Adding Linked Data / Semantic Web information, used to describe relationships between datasets.
- Building APIs around the datasets, to allow developers to get particular parts of the dataset through a standard interface directly from their apps.

Much more detail on the process of opening up data is available in the Open Data Handbook.

Anything else to consider?

You should not open up data containing personal information. If the data can be linked to a particular person, it should be carefully anonymised first before release.

You should consider how often you will update the data, and who will be responsible for doing so. Opening data is not a one-time event; it is a cultural change that requires an ongoing commitment from the organisation as that data changes over time. This can be simplified through various technical means, but will require some thought.

You should provide a way for data users to ask you questions about the data, and answers to any questions that you receive regularly should be added to the public documentation.

Conclusion

We face a crisis in global sustainability that requires urgent action. This requires consensus on the required actions, and innovation to develop new solutions. Open Data can deliver the transparency and accountability to build trust and consensus, and can open up new areas for innovation and improved efficiency. International organisations have a vital role to play in leading the way on sharing data and knowledge.

The process of opening data need not be complex, and does not require all the answers or applications to be known up front. Once data is opened, innovation is enabled both within organisations and in the wider community, creating applications and solutions that could not have been imagined previously.

Further Reading

- [The Open Definition](#)
- [The Open Data Handbook](#)
- [Open Data Commons](#)
- [Linked Open Data: The Essentials](#)
- [Open Data: Driving growth, ingenuity and innovation](#)
- [Open Data from the World Bank](#)

Relevant Organisations

- [The Open Knowledge Foundation](#): promoting Open Knowledge and Open Data across many subject areas.

- [The Open Data Institute](#): Open Data advocacy group set up by the UK government.

- [Climate Code Foundation](#): creating open source climate models.

- [Cleanweb UK](#): promoting innovation on climate change in the UK startup community.

Acknowledgements

Thanks to Velichka Dimitrova of the Open Knowledge Foundation for her valuable feedback on this paper.