Open government data

A guide to best practice in transparency, accountability and civic engagement across the public sector

Transparency
& Accountability
Initiative

c/o Open Society Foundation
4th floor, Cambridge House
100 Cambridge Grove
London, W6 0LE UK

Tel: +44 (0)20 7031 0200
E: contact@transparency-initiative.org

www.transparency-initiative.org

Open government data

Contributor: Centre for Internet and Society – India

Openness in relation to information on governmental functioning is a crucial component of democratic governance. There are few things more abhorrent to democracies than a lack of transparency in their functioning, and secrecy in public affairs is generally a sign of autocratic rule. Such transparency is the foundation for the seeking of accountability from those who exercise power over public policy issues and governmental functioning, including not only governments but also large corporations, trade unions, civil society organisations (CSOs), funding agencies and special interest groups. This information would also include all information on private bodies that can be accessed by public authorities.

Transparency helps citizens to independently evaluate governmental functioning and thus hold accountable any instances of corruption or mismanagement, whether at the level of policy formulation or at the level of implementation. Thus, the freedom of speech and expression and the right to receive information, which are seen as two sides of the same right under most international covenants, are both vitally important in ensuring transparent and accountable governance.

Making public information that is produced by the government is slightly different from merely making public information on governmental functioning. While many instances of the former are subsumed within the latter (e.g. information collected by the government), there are also areas where the two categories do not overlap. Openness with respect to government-produced information is part of the right of the public to access any output of taxpayer funding. Thus the category of 'governmental information' or 'governmental data' can be taken to include information about the government and governmental functioning, as well as information collected and produced by the government.

In addition, there can be two related but independent grounds on which the right of the public to governmental information is often founded. The 'open government data' movement – it is now a demand cutting across multiple nations and deserves to be so called – is predicated upon there being a certain degree of transparency in public functioning, notably through the existence of 'right to information' or 'freedom of information' statutes. Specifically, the open data movement generally understands the public's right to information to include (1) the proactive disclosure of information; (2) the internet being the primary medium for such disclosure; (3) information being made available for access and for re-use free of charge and; (4) information being made available in a machine-readable format to enable computer-based re-use.

As it would be meaningless to demand the additional components that go to make 'open government data' in an environment where the basic right to information does not exist, all recommendations here (including initial steps) presume that such a right exists.

Initial steps

Goal

A commitment by the government to provide proactive disclosure of existing digital data on the web.

Justification

Most governments already rely on computers at least for information storage at most levels even if they often perform information processing and sharing (i.e. conduct governmental transactions whether government-government, government-business, or government-civil society) offline. This information that already exists in a digital form – quite often in the form of text documents and spread sheets – can and should be made public, based on a narrow negative blacklist. This blacklist should have a list of categories of information that should not be made available because of a narrow set of concerns such as privacy and properly classified state secrets. While this will undoubtedly result in the haphazard release of files that may be difficult to comprehend or use effectively, this is not a reason for keeping data offline and out of public reach. Once a process has been initiated of continually putting data up online, the data and the process can themselves be bettered through more elaborate technological and process-related improvements. Proactive disclosure steps can and should be taken even without the implementation of a robust procedural back-end for information gathering, processing and sharing along with the technology that enables it. While such robust information architecture and back-end infrastructure is certainly desirable, it is not necessary for the immediate online release of files that are already in digital format.

Recommendations

1. The government should create a minimal front-facing infrastructure, in terms of both technology (namely, a website) and human resources (people who are tasked with the responsibility of uploading governmental records, documents, reports and other information).

2. A negative list of information that may not be shared should be drawn up by each public authority so that all other material can be made publicly available immediately, keeping in mind the more general guidelines that exist in national and sub-national policies and laws on the right to information.

3. A timeline should be put in place to ensure that proactive disclosure of existing government information continues to happen on a regular basis, until more rigorous steps are taken towards open government data.

More substantial steps

Goal

All government data is made available, in a form that ensures ease of use and reuse.

Justification

Making government data available online is just the first basic step. All information released requires a proper underpinning in informational policy and technological support to realise full transparency, citizen participation and full social and economic value. Governments should use smarter technologies to ensure that the policy commitment to open government data can be realised in practice. In particular, searchability in the system greatly helps to ensure accessibility for persons with disabilities. Such searchability is often easy when it comes to text, but ends up being more complicated in other areas. For this reason, some of the suggestions on this have been kept for the next section (on proposals for most ambitious steps).

Recommendations

Policy and process

1. An information policy should be formulated that deals comprehensively with best practices with regard to information collection, storage, retrieval and management at the national level, and that allows for the adoption of that policy either with modification or directly by sub-national governments.

   a. Part of this policy must ensure that most new information is either created in a digital form, or is digitised from paper as soon as is practicable, and that later transactions of this information happen, as far as possible, over electronic modes of communication.

   b. This policy must also ensure that as much as electronic receipt of governmental information is seen as a right of citizens, so is non-electronic receipt.

2. A technological policy should be formulated that mandates the use of open standards in all e-governance to promote interoperability and prevent vendor lock-in, with only temporary and limited exceptions.

   a. This must be accompanied by a document on technological architecture (whether called an e-governance interoperability framework (e-GIF) policy, or a national enterprise architecture (NEA)) that lays down the broad parameters of the technology framework to enable the information architecture policy, including the metadata standards.

3. The ability to reuse the published data must be guaranteed as part of a public sector information/open government data policy. This is crucial to enable journalists, CSOs and others.

4. All information must be provided free of cost at least in cases where:

   - The government is not monetising the data, nor has plans to do so; or
   - The data is for use by individuals and small and medium enterprises; or
   - The data is available without any special fees under right to information/freedom of information statutes.

Recommendations

Technology

1. All public authorities must be made to ensure that they use open standards, such as Unicode, prescribed in the e-GIF/NEA. In addition, their data processing and publishing processes must comply with those laid out in these architectural documents.

2. Sector-specific and use-specific metadata must be included in all files and objects made available to the public, so that when they use the services to retrieve objects they can make sense of the objects and manipulate them appropriately.

3. This metadata must be standardised, as this is a crucial requirement to enable easy categorisation and searching of information. An important part of searching through the data is also searching through the full contents of the datasets.

Most ambitious steps

Goal

To translate the publishing of open governmental data into better data via input from the public.

Justification

Public outreach and citizen-oriented tools are crucial to ensuring a vibrant online and offline public sphere where government data are used and discussed and a feedback loop is created, rather than this being a mere data dump. Using service-oriented architecture will help in ensuring platform independence, better scalability, greater code reuse, higher availability of services, parallel development of different components and many other benefits in terms of provision of data for governments. A robust service-oriented architecture will enable citizens to be treated as yet another client asking for information, and will enable useful application programming interfaces (APIs) to be built that will allow for easy access for power users to the data.

Integration with social media is a must, because it allows governments to reach vast networks of people at once and defray costs. Such integration will allow governments to go where many citizens are, rather than trying to get the citizens to come to them. However, care must be taken to ensure that such integration is done with adequate safeguards for privacy, long-term archival capability and data portability.

Recommendations

Policy and process

1. The pro-elite bias that is often inherent in online technologies must be actively neutralised through policy. Such a policy must be designed to ensure that there is no elitist capture of the benefits of open government data, and that there is active promotion of 'offline translation' of data, especially in technologically divided countries where the gap between those who have access to technology and those who do not is wide.

2. Governments must allow for correction of data by the public.

3. Offline translation of data must be facilitated, especially in technologically poorer countries.

Technology

1. Documents should be structured with semantic mark-up, which allows for intelligent querying of the content of the document itself. Before settling upon a domestic usage-specific semantic mark-up schema, well-established XML schemas should be examined for their suitability and should be used wherever appropriate.

2. Multiple forms of access must be provided to the data, and it must be made available interactively through the web for non-technical users. For more advanced users, the data must be available for bulk downloads, and it should also be accessible through well-documented open APIs.

3. There should be a single-point portal (similar to the UK's Data.gov.uk) to provide access to different public authorities' data.

4. All data should be Cloud-based to the extent that it ensures lower overheads for the government.