



Open Government Data Strategy

www.data.gov.bd

Statistics and Informatics Division
Ministry of Planning

Index

1	Introduction	
1.1	Vision	
1.2	Mission	
1.3	Objective	
1.4	Concept of open data	
1.5	Benefits of open data	
2	Basic Principles of Open Data	
3	Basics Standards of Open Data	
4	Strategy	
4.1	Embedding open data practices	
4.2	Identifying candidate open data	
4.3	Assessing open data candidates	
4.4	Releasing open data	
4.5	Publishing open data	
4.6	Publishing case studies	
4.7	Raising staff awareness about open data	
4.8	Engaging with the open data community	
4.9	Engaging through the Open Data Portal	
4.10	Engaging through open data events	
4.11	Engaging with existing stakeholder groups	
4.12	Engaging with new interest groups	
4.13	Enhancing our open data based on demand	
4.14	Content Management	
4.15	Control and Ensuring success	
5	Open Government Data Working Group	
6	Open Data Executive Team	
7	Open Data Policy	
8	Data Portal	
9	Conclusion	

Introduction

Technology and data have taken the driving seats for changing the world we live in. New groundbreaking uses of data, novel insights into existing data and innovative ways of delivering that data to citizens via mobile devices and wearable technology all have the potential to improve public services such as health, education, environment, planning and many more. In connection with Digital Bangladesh mission, to ensure services at citizens' doorsteps, Government of Bangladesh creates and maintains vast amounts of many different types of data in different sectors including Health, Education, Agriculture, Tourism, Finance & banking etc. Opening up access to this type of data to citizens and the private sectors has been demonstrated to have many positive impacts such as increased public service efficiency, accountability of government and empowering citizens. Open data can support reform and improvement through greater transparency and citizen participation. Opening up this data to the public will help increase the usability of the data through people's taking, transforming and adding value to the data. The release of public sector data in open data formats has been demonstrated to be a driver in terms of growth, job creation, research and innovation which itself has been identified as one of the primary drivers of economic growth.

1.1 Vision Statement

Data for all

1.2 Mission Statement

- Forming open government data principles by means of innovation in data management and change in tradition for information
- Ensuring public service availability efficiently, promoting innovation & research and improving national economy in Bangladesh.
- Provisioning structured way of access to information for mainstreaming public inclusions in national economy through open data flow among individuals and agencies.
- Establishing good governance by ensuring transparency & accountability through citizens' participation

1.3 Objectives

- To encourage developing innovative solutions for better public service delivery
- To enhance scope of research to identify and develop innovative solutions.
- To create opportunities for new jobs and more investment.
- To make government more transparent and accountable

1.4 Concept of open data

Open Data is the idea that certain data should be made available to everyone to use and republish as they wish, with no or very limited restrictions from copyright, patents or other controls. This data or content is released in such a format that allows anyone to access, freely use or reuse, distribute or re-distribute and share without any royalties or restriction either by technologies or by legal constraints. Attributes of Open Data need to be share-alike.

Open data does not mean that a government or other entity releases all of its data to the public. It would be unconscionable for the government to give out all of your private, personal data to anyone who asks

for it. Rather, *open data* means that whatever data is released is done so in a specific way to allow the public to access it without having to pay fees or be unfairly restricted in its use.

Open data, though straightforward in principle, requires a specific approach based on the agency or organization releasing it, the kind of data being released and, mostly the targeted audience.

There are two important elements to openness:

- Legal openness: General access needs to be allowed to get the data legally, to build on it, and to share it. Legal openness is usually provided by applying an appropriate Open License System which allows for free access to and reuse of the data, or by placing data into the public domain.
- Technical openness: There should be no technical barriers to use the data. For example, providing data as printouts on paper (or as tables in PDF documents) makes the information extremely difficult to work with.

1.5 Benefits of Open Data

a. Benefit to government

- Increased tax revenues through increased economic activity
- Creation of jobs
- Reduction in data transaction costs
- Increased service efficiency (especially through linked data)
- Increased GDP
- Encouraged entrepreneurship (economic growth)
- Better decision-making based on accurate information

b. Benefit to private sector

- New business opportunities for services / goods
- Reduced costs for data conversion (no need to convert into raw formats anymore)
- Better decision-making based on accurate information
- Better-skilled workforce

c. Benefit to NGOs / civil society

- Better informed monitoring
- New avenues for project action: building tools/applications
- Increased sustainability potential through increased capacity

d. Benefits for individual users:

- Easier interaction with the government (eg. crime reporting, potholes reporting, fix-my-street).
-
- Better navigation facilitated by mapping data, databases of points of interest, etc. (eg. route planning, public transportation schedules).
- Improved data availability to the researchers in an efficient way on emerging issues.

e. General Benefits:

- Open data empowering transformation in specific sectors such as the financial one;
- Open data generating new kinds of Public-Private partnership models;
- Open data policies accelerating the process of private businesses releasing its own data;
- Open data disrupting traditional business models, lowering entry barriers and making the services industry more modular.
- Better availability of data to the researchers and beneficiary.

f. data sharing policy:

- Maximizing Use: Ready access to government owned data will enable more extensive use of a valuable public resource for the benefit of the community.
- Avoiding duplication: By sharing data the need for separate bodies to collect the same data will be avoided resulting in significant cost savings in data collection.
- Maximized integration: By adopting common standards for the collection and transfer of data, integration of individual data sets may be feasible.
- Ownership information: The identification of owners for the principal data sets provide information of prioritized data collection programs and development of data standards.
- Better decision-making: Data and information facilitates making important decisions without incurring repetitive costs. Ready access to existing valuable data is essential for many decision making tasks such as protecting the environment, development planning, managing assets, improving living conditions, national security and controlling disasters.

1. Basics of Open Data Principles

Government data shall be considered open if it is made public in a way that complies with the principles below:

a. Complete

All public data should be made available. Public data is data that is not subject to valid privacy, security or privilege limitations. While non-electronic information resources, such as physical artifacts, are not subject to the Open Government Data principles, it will Always be encouraged that such resources be made available electronically to the extent feasible.

“Bulk data” means that an entire dataset can be acquired. Even the simplest of applications, such as computing the sum of line items, requires access to the entire dataset. This principles principle implies that bulk data should be made available *before* “API”s are created because “API”s typically only return small slices of the whole data.

b. Primary

Data should be as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.

If an entity chooses to transform data by aggregation or transcoding for use on an Internet site built for end users, it still has an obligation to make the full-resolution information available in bulk for others to build their own sites with and to preserve the data for posterity.

c. Timely

Data should be made available as quickly as necessary to preserve the value of the data.

d. Accessible

Data should be available to the widest range of users for the widest range of purposes. Data need to be made available on the Internet so as to accommodate the widest practical range of users and uses. This means considering how choices in data preparation and publication affect access to the disabled and how it may impact users of a variety of software and hardware platforms. Data must be published with current industry standard protocols and formats, as well as alternative protocols and formats when industry standards impose burdens on wide reuse of the data. Data should be

accessible not only through navigating web forms but also by automated tools without having technological restrictions.

e. Machine processable

Data should reasonably be structured to allow automated processing. The ability for data to be widely used requires that the data be properly encoded. Free-form text is not a substitute for tabular and normalized records. Images of text are not a substitute for the text itself. Sufficient documentation on the data format and meanings of normalized data items must be available to users of the data. Data published by the government should be in formats and approaches that promote analysis and reuse of that data.” The most critical value of open government data comes from the public’s ability to carry out its own analyses of raw data, rather than relying on a government’s own analysis.

f. Non-discriminatory

Data needs to be available to anyone, with no requirement of registration. Anonymous access to the data must be allowed for public data, including access through anonymous proxies.

g. Non-proprietary

Data needs to be available in a format over which no entity has exclusive control. Proprietary formats add unnecessary restrictions over who can use the data, how it can be used and shared, and whether the data will be usable in the future. While some proprietary formats are nearly ubiquitous, it is nevertheless not acceptable to use only proprietary formats. Likewise, the relevant non-proprietary formats may not reach a wide audience. In these cases, it may be necessary to make the data available in multiple formats.

h. License-free

Data should not have dependencies like copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed. Government information is a mix of public records, personal information, copyrighted work, and other non-open data, it is important to be clear about what data is available and what licensing, terms of service, and legal restrictions apply. Data for which no restrictions apply should be marked clearly as being in the public domain. Requiring attribution to the government, even though attribution might be reasonable in other contexts, would constitute a major policy shift in the country with significant legal implications for the press.

3. Basics of Open Data Standards

Web-services technology has placed in the mainstream the idea of applications working together as single distributed system across the Internet. Many statistical standards are beginning to use this type of architecture: in DDI (Data Documentation Initiative), the idea of having a centralized registry to function as a "question bank" has emerged; in SDMX (Statistical Data and Meta Data Exchange), a set of standard interfaces have been provided for interacting with a registry that gives visibility into sets of aggregated data, repositories of related structural and descriptive meta data, and the process-related information about the use and exchange of this data and meta data. These ideas are very similar, despite being applied to different parts of the statistical life cycle.

The application of registries and service-oriented architectures to the statistical realm is an idea that is coming of age, but there is a risk that different standards and specialties within the statistical realm will apply this technology in non-inter operable ways. This is also, however, an opportunity for the ODaF to help design and promote a single aligned architecture, and to avoid the problems inherent in having multiple competing standards. These data standards should be compliant with Bangladesh National Enterprise Architecture (NEA).

3.1 Data and Metadata Models:

There are today many standard modeling techniques which promote interoperability across standards. Standards of Bangladesh Open Data and Metadata model will include the SDMX Information Model (a meta-model for aggregated statistical data and related metadata); ISO/IEC-11179 and the Common Metadata Repository (CMR) extensions (for defining semantics and the collection and management of data and metadata); ISO-15000 Core Components Technical Specification (for modeling transactional data); and the Neuchatel model for classifications schemes. There are many other useful models as well, particularly the model emerging from DDI in their next release for survey instruments, raw data, micro data, aggregate data and related metadata. Many of these standards are already drawing on one another, and a full alignment is possible. Such models provide a basis for the mapping of other models into a coherent whole using a standard semantic metamodel such as ISO-11179 as a central “pivot”.

3.2 Data and Metadata Formats:

Many standards provide formats for statistical data and metadata. Many of these - but not all - are described using the XML standards; some are in EDIFACT syntax (GESMES) or in proprietary formats of various types. Many of these formats are based on the various meta-models described above, and are also to some extent aligned. By formally mapping them, it is possible that a set of interoperable applications can be created, facilitating the exchange and use of statistical data and metadata. Standards of Bangladesh Open Data and metadata format will follow the above mentioned formats.

3.3 Statistical Registries:

With the advent of service-oriented architectures, the web-services concept of a "registry" becomes important. Bangladesh Open Data can include two horizontal technology standards dealing with registries: ISO-1500, which describes the ebXML Registry/Repository, and OASIS' UDDI registry specification. These generic registry standards have been refined in various ways for use within statistical applications, most notably in version 2.0 of the SDMX Technical Specifications. Having a standard set of registry interfaces based on a standard registry model is the lynch-pin to having a universal statistical architecture. ODaF [Open Data Foundation] policies/guidelines may be followed in this regards.

As the use of standard registries develops, the ODaF [Open Data Foundation] is well-positioned to recommend how it can be employed to support the various standards within the statistical life cycle and to facilitate the development of tools to make the registry-based vision a reality.

3.4 Semantic Registries:

ISO/IEC-11179 provides the concept of semantic registries, where the meaning of data and metadata elements can be effectively managed. Alignment with this formal definition of meaning, in an accessible fashion, is critical to the ODaF's [Open Data Foundation] vision of interoperability, as an aligned part of an overall statistical architecture.

3.5 Metadata Repositories:

Today, there is a growing emphasis on the use of public metadata repositories to support data quality initiatives. This has been a major driver in the development of the SDMX standards, which themselves drew on many existing initiatives. Other standard views of metadata repositories also exist, notably the CMR extensions to ISO/IEC-11179 and metadata repositories based on DDI, made available by many data archives. In the larger world of digital libraries, many useful standards such as Dublin Core (among others) can be employed to describe collections of digital artifacts. These repositories form a critical part of a universal statistical architecture, containing much of the valuable metadata which is today difficult to discover and access. By facilitating the alignment of metadata standards, and providing tools to work with them, metadata repositories will more easily become a part of an overall statistical network. Strategically, Bangladesh Open Data can have such Metadata repository standards.

3.5 Data Types

Data can be of Unstructured, Structured and Database in nature. Unstructured data includes electronic mail messages, scanned memoranda and documents, images, and video of, for example, car stops performed by a police department.

But considerable amounts of the bulk data (large datasets) are structured. Structured data can be placed on an open data site easily. The simplest way to visualize structured data is as a table or a spreadsheet with rows and columns. Each column contains specific data, e.g. a street address or a telephone number. Each row is another instance of the data.

A third type of data is a database. Often a database consists of many different tables, each of which is structured, but each of which probably makes little sense unless joined or combined with data from another table. An example of such a database is a human resource management system (HRMS). One table might contain employee name and personal characteristics such as date of birth, social security number, age, employee number and so forth. A second table in the database might have names of dependents of all employees, plus their personal characteristics. A third table might contain all the training classes and certifications the employee has taken or earned. Databases present a special problem in open data, in that the data to be “opened” often must come from a query or report and then may need to be redacted to remove restricted information such as Credit Card Number.

4. Strategy

Data which can be made open will come from a variety of sources including registers, databases, spreadsheets, Censuses, surveys and geospatial datasets. This data is collected and managed by the public sector body to carry out its duty or public task. Text-based data such as day to day emails, memos, businesses cases and reports which are generated from aggregating and manipulating raw data (which can be released) are excluded.

The public sector holds vast amounts of data and although the aspiration of this strategy is that the release of data becomes part of the data management cycle, this will take time to embed and therefore we need to manage expectations. In order to get the balance right we will priorities what is released in relation to external demand and interest with the acknowledgement that some datasets will not be able to be released due to the exceptions as mentioned in the principles. Initial balancing will have to be done to prioritize what data is published prior to reaching ‘public by default’. Priority however will be given to datasets that are deemed to be of “high value” in terms of their potential for reuse.

Open Data Strategy aims to:

- Embed open data practices within the Government.
- Engage with the open data community.
- Enhance open data based on demand.

Success will be attained through strong management, good governance, performance measurement and continuity plan.

4.1 Embedding open data practices

Bangladesh Government can continue to embed open data practices both within its existing processes and in re-engineered processes. Coordination of open data activities can be done centrally to ensure a consistent approach, knowledge sharing and efficient use of resources.

4.2 Identifying candidate open data

Government will identify data to be released as open data in a number of ways:

- Review data already published on National Portal
- Review requests for datasets received directly from the public through departmental correspondence process.
- Review Data Searching Trend in National Portal and in e-Thathyakosh.
- Data request Survey by Open Data Portal
- Engage Open Data Working Group to be formed by the government to assess existing data repository both in Public and Private Sector.
- Engage with potential users of open data to understand the demand for datasets.

4.3 Assessing open data candidates

Data will be assessed to determine if it can be published as open data. Private and commercially sensitive data will be protected and will only be released in summary or de-identifiable formats. Data custodians will assess whether the data is releasable or not. The assessment will be validated by competent authority of the Government under some standards and policies. Assessments may also capture information to help determine the sequence of the release of data. This may include:

- The potential value achieved via the proposed use of the data.
- The public demand for the dataset.
- The complexity of converting the data to open data.

4.4 Releasing open data

Releases of open data will be scheduled based on assessment results, logical groupings of datasets and available resources. Where practical, quality checks will be performed on open data prior to release.

4.5 Publishing open data

Open data will be cataloged and published so it can be easily discovered in the Government Open Data Portal integrated to Bangladesh National Portal. Government can have Open Data Blog as the online network for Open Data Working Group.

4.6 Publishing case studies

Case studies will be cataloged and published so they can easily be accessed through the Government Open Data Portal integrated to Bangladesh National Portal.

4.7 Raising staff awareness about open data

Open data concept and practice will be communicated to departmental staff and private sector focal persons via Intranet, National Portal and Departmental messages. Open Data Governance Group and Open Data Executive Team will be established to champion the open data awareness within the Government. Innovation Officers will represent the initiative in different layers of Public Service Delivery Channel (PSDC).

4.8 Engaging with the open data community

Government of Bangladesh recognizes that active engagement with the open data community is the key to the success of the open data initiative. Without an active community using open data, the release of data is ineffective. The Government will promote the use of open data through a variety of social media platforms, such as Facebook, Twitter, etc.

4.9 Engaging through the Open Data Portal

Requests for new data and improvements to data will continue to be received via Government Open Data Portal. All questions and responses will be encouraged and coordinated centrally.

4.10 Engaging through open data events

Engagement with users of open data will be ensured via Government open data initiative by participating in organized events, working groups and competitions.

4.11 Engaging with existing stakeholder groups

The Government will include open data in discussions with a number of existing stakeholder groups to better understand their requirements. Government will also support the Open Data Working Group in its interactions with various stakeholder groups.

4.12 Engaging with new interest groups

Communities of interest can be formed basing on various types of departmental data and more direct engagement may be required. Government of Bangladesh will encourage participation in conversations with these new interest groups on data release priorities, preferred data formats, helping understand the data, receiving feedback on data quality and capturing case studies.

4.13 Enhancing open data based on demand

Enhancements to open data will be demand driven and prioritized based on value to the people of Bangladesh. These enhancements may take many forms including:

- frequency of publication;
- new data download formats;
- correcting reported errors and
- Using standards-based classification schemes.

4.14 Content management

Government of Bangladesh will have a guideline in place which will contain necessary principles of developing contents, adding data to public portal in compliance with state laws and social sensitivities. Both technology controls and civil security controls will clearly be defined in the guideline but it will not be an obstacle for freedom of openness until and unless any content is not harmful to any citizen or to the state anyway. This content delivery channel will also be interoperable. It will also be a national data/content interchange backbone with all necessary information security controls in place.

4.15 Control and Ensuring success

The Open Data Strategy will be governed by an Open Data Executive Team, assisted by the Open Data Working Group. Government may appoint or nominate someone who will be accountable for ensuring the objectives. The strategy will be reviewed every year with progress reports being provided periodically.

5. Open Government Data Working Group

The purposes of the Working Group are to:

- act as a central point of reference for people who are interested in Open Government Data;
- develop principles for making official information legally and technically open;
- document the background and status of initiatives to make official information open;
- support the development of open government data catalogs, and ensure technical interoperability assessment and
- Support Open Data Executive team recommending change in policies or adding new policies for Open Data Initiative.

Government will form this Group under Statistics and Informatics Division (SID) comprising of Government officials and private sectors related interested professionals. This will be an open working group with broader representation from civil society, ICT Industries, Development partners, Academicians, Students, Researchers and relevant professionals.

6. Open Data Executive Team

An Executive Team will lead Open Data initiative. This team will consist of renowned industry Experts who have experience in data management and compliance solutions, Domain experts who have consistent track record of innovation and customer focus, representatives of government agencies who deal with robust data systems and associated regulatory systems.

7. Open Data Policy

Government of Bangladesh will have a policy in place to govern, regulate and manage Open Data System, its compliance and all of its related activities in the country. Open Data Working Group will assist the Open Data Executive Team to formulate such policy and will be authorized

By competent agency of the Government

8. Data Portal

Government of Bangladesh will have a Data Portal as an integral part of National Portal. This will provide an easy way to find access and reuse public datasets. Request for new datasets can also be submitted using that portal. This portal will have an open Data Catalog mentioning classified data attributes along with data owner and data custodian. This portal will publish open data statistics like discoverable datasets, API enable resources, data groups etc.

9. Conclusion

Through legitimate business processes government generates significant amounts of data about people, property, licenses, crimes, public health and a wide variety of other entities. This data is used by elected and senior officials to make laws, set policies and operate government services. As far as good governance is concerned and considering e-democracy, citizens must be engaged and involved in the processes of law and policy making, and even in service provision through the digital information backbone.

The government open data movement endeavors to open as much of this data as possible for use by citizens, advocacy groups, students, researchers and private businesses, within the constraints of protection of privacy and security. A small but significant number of governments embrace the concept of open data and have established open data portals on the Internet.

At the same time, the Internet is widely used by consumers, citizens and businesses for social media, financial transactions, purchases and simply to find information. Businesses, especially Internet-based ones such as Amazon and Google, are collecting vast amounts of data about their users. Some of this data is open, and most of it is bought-and-sold on private exchanges.

A number of new businesses can use government open data along with other publicly and privately available information. These businesses can be able to improve quality of life by producing new products and services. Citizen advocacy group can use open data to produce new insights into public policy problems. These insights will improve the process of creating laws and public policy, and improve the delivery of government services.

There can have budgetary and cultural constraints. But senior elected officials and leaders like MP and local government leader can follow a play book with specific actions to overcome the obstacles and open their government data for use by citizens, academics and private businesses. The result will be improved and wider citizen engagement with government and will improve quality of life for people.

Acronyms

CMR: Common Metadata Repository

DDI: Data Documentation Initiative

ebXML: e-Business XML

EDIFACT: Electronic Data Interchange for Administration, Commerce and Transport

GESMES: Generic Statistical Message for Time Series

ODaF: Open Data Foundation

OASIS: Organization for the Advancement of Structured Information Standards

SDMX: Statistical Data and Metadata Exchange

UDDI: *Universal Description, Discovery and Integration*

XML: Extensible Markup Language

PSDC: Public Service Delivery Channel.