University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal) | Libraries at University of Nebraska-Lincoln

April 2018

Evaluation of Open Data Government Sites: A Comparative Study

JAYANTA KR NAYEK
Dr.B.R.Ambedkar College, Betai, Nadia, West Bengal, India, j.nayek@gmail.com

Follow this and additional works at: https://digitalcommons.unl.edu/libphilprac

 Part of the Library and Information Science Commons

NAYEK, JAYANTA KR, "Evaluation of Open Data Government Sites: A Comparative Study" (2018). *Library Philosophy and Practice (e-journal)*. 1781.
https://digitalcommons.unl.edu/libphilprac/1781

Evaluation of Open Data Government Sites: A Comparative Study

Jayanta Kr. Nayek
Librarian, Dr.B.R.Ambedkar College, Betai, Nadia
West Bengal, India
Email: j.nayek@gmail.com

Abstract

The problems of open data government sites are mainly in the way data is presented by various government data repositories. There are no metadata standards, specifications, or protocols to achieve better discoverability and interoperability. This paper aims to find a way to solve the problems with open data government sites and propose a framework to provide web based data services which will be semantically structured and also propose a common metadata standard or mechanism for interoperability. Six data government sites have been selected for the study, according to the more number of dataset available and also to cover all continents of the world (followed convenience sampling logic). A framework has been suggested in this paper and it is expected that some government sites may follow the framework in the near future.

Keywords: Open Government Data (OGD), interoperability, DCAT (Data Catalog Vocabulary), metadata, data provider, service provider, comparative study.

1. Introduction

The term "open data" is a recent buzzword, getting popularity with the progress of the World Wide Web and specifically, with the drive of Open Government Data (OGD) initiatives such as data.gov.uk, data.gov.in, data.gov etc. Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright or any other restrictions (Ubaldi, B. (2013))[19]. The OGD initiative facilitates publicly available government data to be freely available practicing open data and open development protocols.

After the Budapest Open Access Initiative (2002) and the Open Access movement governments have started to look into the prospects of providing open access to their data repositories. The advantages of OGD is that it enables greater government efficiency through an information infrastructure that allows for better data re-use within the public sectors and inter-agency coordination.

But the problems with data government sites is mainly the way data is presented by various government data repositories and no standards or specification or protocols are in place to achieve interoperability. These problems led to develop a framework to provide web based data services which will be semantically structured and also propose a common metadata standard or mechanism for interoperability.

1.1 Background

The development of Open Government Data (OGD) initiatives/Open Government Data portals have started in mid 2000s. The numbers of portals are growing rapidly. The reason behind it is that government data are becoming more easily accessible and be used for various other purposes. Another reason may be that open government data is expected to improve the decision making for both the government and the public. OGD can be used to help the public better understand what the government does and how well it performs, to hold it accountable for unfinished/unachieved results. It also helps to generates insights into how to improve government performance. It is important for governments to seek feedback from the public on the usefulness, relevance and accessibility of their data, in order to allow for continuous improvement.

1.2 Definition of Open Government Data[17]

According to Open Government Data[1] website, "open" means data is open ie. free for anyone to use, re-use and re-distribute and "open government data" means data and information produced or commissioned by government or government controlled entities. The government data shall be considered open if it is made public in a way that complies with the principles[16]: i) Complete ii) Primary iii) Timely iv) Accessible v) Machine processable vi) Non-discriminatory vii) Non-proprietary, viii) License-free

1.3 Scope of the study

As, it is already mentioned that the government data shall be considered open if it is made public in a way that follows the above mentioned eight (8) principals. Now if we consider the fourth and fifth principal i.e. related to accessibility (Data is available to the widest range of users for the widest range of purposes) and machine processable (Data is reasonably structured to allow automated processing). The data may not be useful for widest range of purposes because there is no metadata standard for most of the data available on government sites, so interoperability is a problem.

Now if we consider the fifth principal, the data may not allow automated processing because for automated processing it should be in well structured form. Though data is available is in different and varied formats in different formats in different data government site. So, I could propose a metadata standard (metadata promises discoverability which will facilitate Linked Open Data) that have been used to improve or achieve interoperability among metadata schemas for the purposes of facilitating conversion and exchange of metadata and enabling cross-domain metadata harvesting[3] and it would work better to overcome the problems with the data government sites.

For this purpose, I have selected six data government sites for the study, according to the more number of dataset available and also to cover all continents of the world (followed convenience sampling logic) and use two methodologies a) overall assessment of the status of OGD portals of different national governments and b) review of six selected cases of OGD portals.

1.4 Objective of the study

The objective of the study is to obtain a way to solve the problems with data government sites and propose a framework to provide web based data services which will semantically structured and also propose a metadata standard, which is based on Dublin Core and DCAT (Data Catalog Vocabulary) to achieve interoperability. To make the data discoverable, we need to expose our data through OAI-PMH/OAI-ORE protocols.

1.5 Literature Review

According to Ubaldi (2013)[19], a number of challenges may be associated with the implementation of OGD initiatives which, if not properly tackled, might obstruct or restrict the capture of benefits of national efforts aimed at spurring OGD. The problems are:

a) Government data are often un-harmonised as every public agency has its own set of data, formats and standards. This can make it difficult from the user perspective to know which piece of data is valid or should be trusted.

b) Interoperability remains an unresolved issue in e-government, and can potentially have an impact on OGD development as well. Dealing with OGD in general, and open data file formats in particular, can facilitate IT system interoperability in government open data projects. Interoperability is a major concern for policy makers working on the implementation of OGD.

According to Nugroho (2013)[15], in general, there is a lack in guidelines to regulate and help the process of opening data. Many countries are in different stages in developing these guidelines. A

field of study that is lacking is about how countries can learn from each other in developing the necessary guidelines.

According to Braunschweiget. al. (2012)[2], just publishing the data on the web is not enough. To truly advance the open society, the publication platforms need to fulfill certain legal, administrative as well as technical requirements.

1.6 Features of Open Government Data Sites

Here I have enlisted some of the features of government data sites:

i) The dataset is readily and uniformly accessible.

ii) Anyone can read the data but also perform more advanced operations such as searching and filtering.

iii) One can combine datasets with other web services to create new mashups and applications.

iv) The datasets are available in different formats and it can be downloaded easily.

v) It is a place to manage public/non-public datasets: create new entries, modify existing ones, and delete any datasets as needed.

vi) It is a platform for single-point access to datasets and applications published by Ministries/Departments/Organisations of the Government.

1.7 Observations from Open Government Data Sites

The following observations are made after a study of few governments' open data sites:

i) Mostly data sets are in structured format (e.g. XML, CSV, XSL, JSON etc).

ii) The files contain structured data.

iii) The focus of data being published does not correlate with the data that most viewed by users.

iv) Most of them provide metadata but they are not in a structured format.

v) Different metadata standards are followed by different sites.

vi) The search result is not based on semantic web philosophy. The search results are mere tables but not answers to the exact queries.

vii) There appears to be no mechanism for intelligent agents to automatically collect data or metadata as in case of digital repositories of publications where OAI-PMH or OAI-ORE are used to make the data to be harvested by any service provider.

2. Different "data gov" Sites

The list of countries offering easy to find, download or access open data sets continues to grow. According to Open Data Site Finder, there are at least fifty countries with two hundred and ninety seven sites[20]. Here is a list of the most useful government open data sites around the world:



Figure1: Open Data Site Finder
Source: http://dataremixed.com/2013/08/worldwide-open-data-sites/

Australia (data.gov.au), Brazil (dados.gov.br), Canada (data.gc.ca), France (data.gouv.fr), Germany (govdata.de), India (data.gov.in), Italy (dati.gov.it), Kenya (Opendata.go.ke), New Zealand (data.govt.nz), Spain (dato.gob.es), Switzerland (opendata.admin.ch), United Kingdom (data.gov.uk), United States of America (data.gov), etc. Out of which, I have selected six data government sites for the study, to cover all continents of the world.

2.1 Data.gov.in (India)[7]

Open Government Data (OGD) Platform India (data.gov.in) is a platform that supports Open Data initiative of the Government of India. The portal is expected to be used by Government of India Ministries/Departments, their organizations in order to publish the collected datasets, documents, services, tools and applications for public use. It aims to increase transparency in the functioning of Government. It is also expected to open avenues for many more innovative uses of Government Data to convey diverse viewpoints.

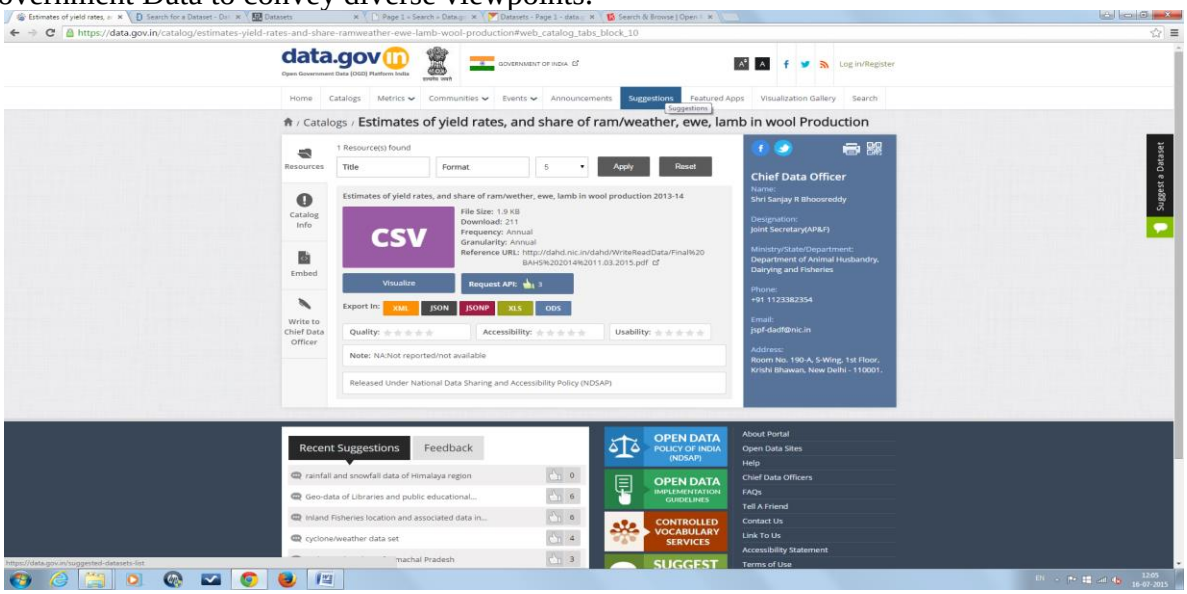


Figure 2: Screenshot of data.gov.in

2.2 Data.gov (USA)[5]

Data.gov is the home of the U.S. Government's open data. One can find Federal, state and local data, tools, and resources to conduct research, build apps, design data visualizations, and more. The Data.gov team works at the U.S. General Services Administration and data on the site are provided by hundreds of organizations, including Federal agencies.

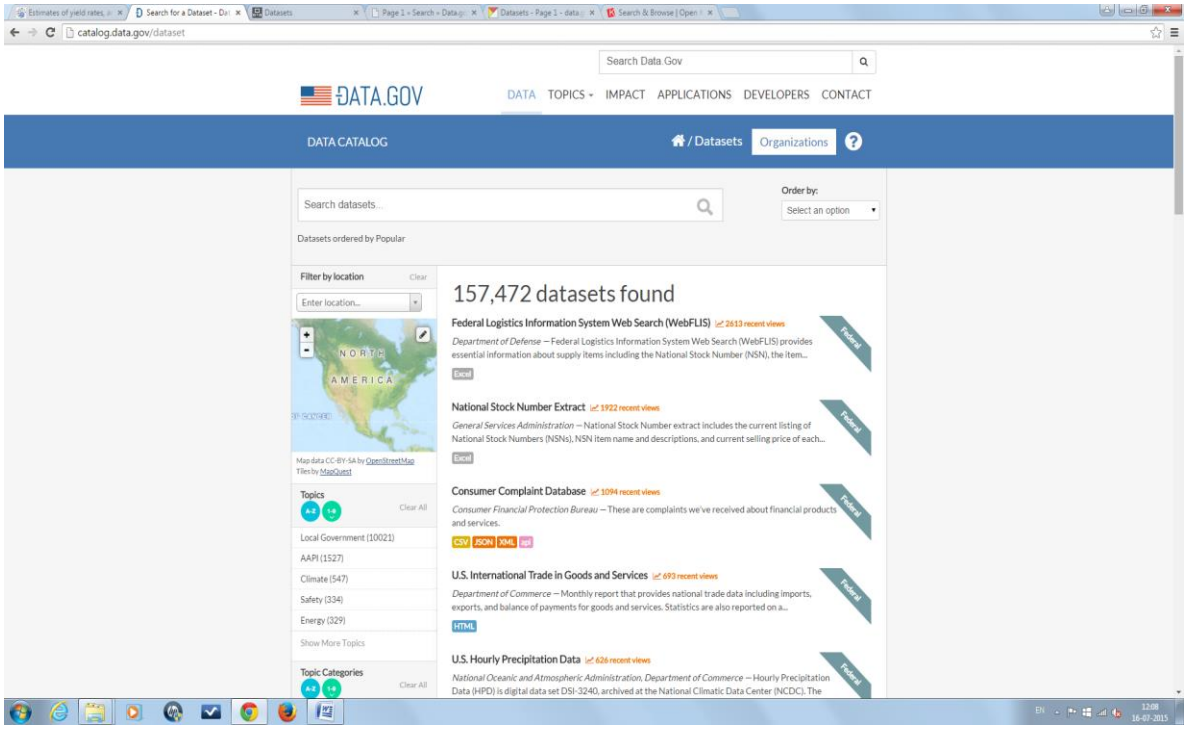


Figure 3: Screenshot of data.gov

2.3 Data.gov.au (Australia)[6]

Data.gov.au provides an easy way to find access and reuse public datasets from government. The main purpose of the site is to encourage public access to and reuse of government data by providing it in useful formats under open licences. The purpose of this online service is to encourage public access to and reuse of government data by providing it in useful formats and under open licences. The site provides both downloadable datasets and links to online data services provided by other government sources. Improving the quantity and quality of the site's data will be an ongoing process. The datasets provided through data.gov.au have been created by many different government agencies.

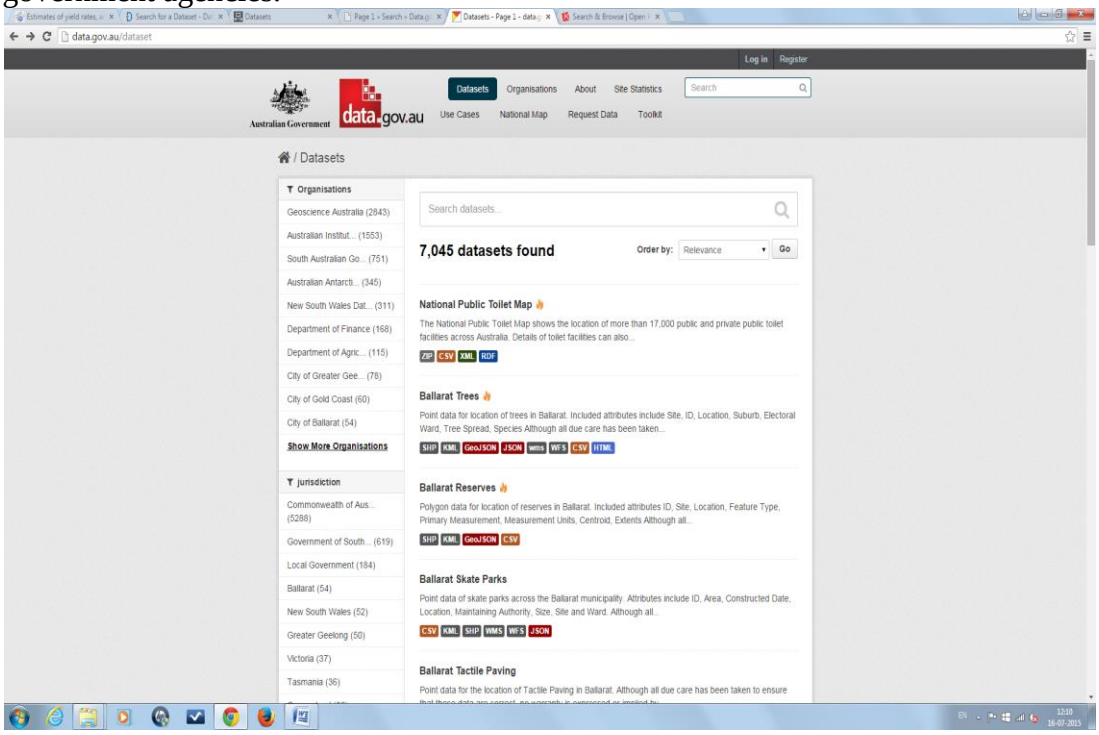


Figure 4: Screenshot of data.gov.au

2.4 Data.govt.nz (New Zealand)[9]

The Open Data Catalogue was launched 5 June 2009 as the site to locate government data on the internet. The aims of the site are to:

i) List all of the datasets available to members of the public.
ii) Provide a place for people to comment on the datasets.
iii) Make it easy for people to find the information they are after and who they need to contact.
iv) Provide a voice for the data using community, both professional and casual.

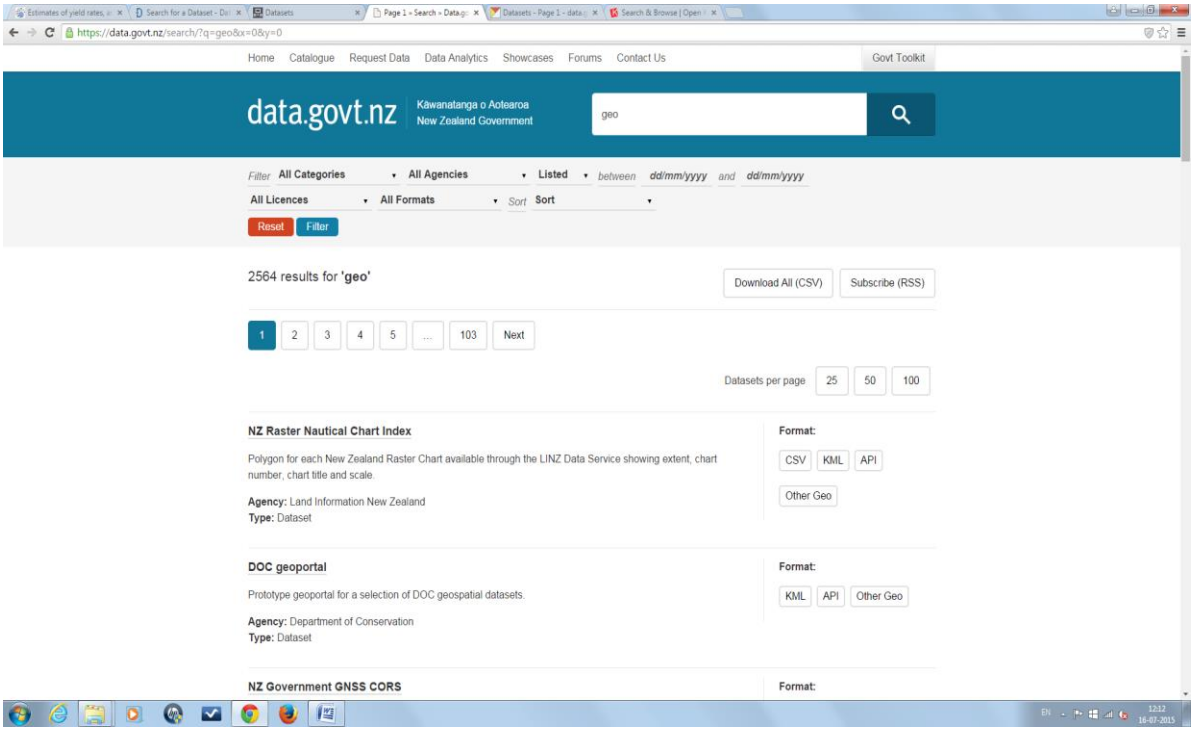


Figure 5: Screenshot of data.gov.nz

2.5 Data.gov.uk (UK)[8]

The government is releasing public data to help people understand how government works and how policies are made. Some of this data is already available, but data.gov.uk brings it together in one searchable website. Making this data easily available means it will be easier for people to make decisions and suggestions about government policies based on detailed information.

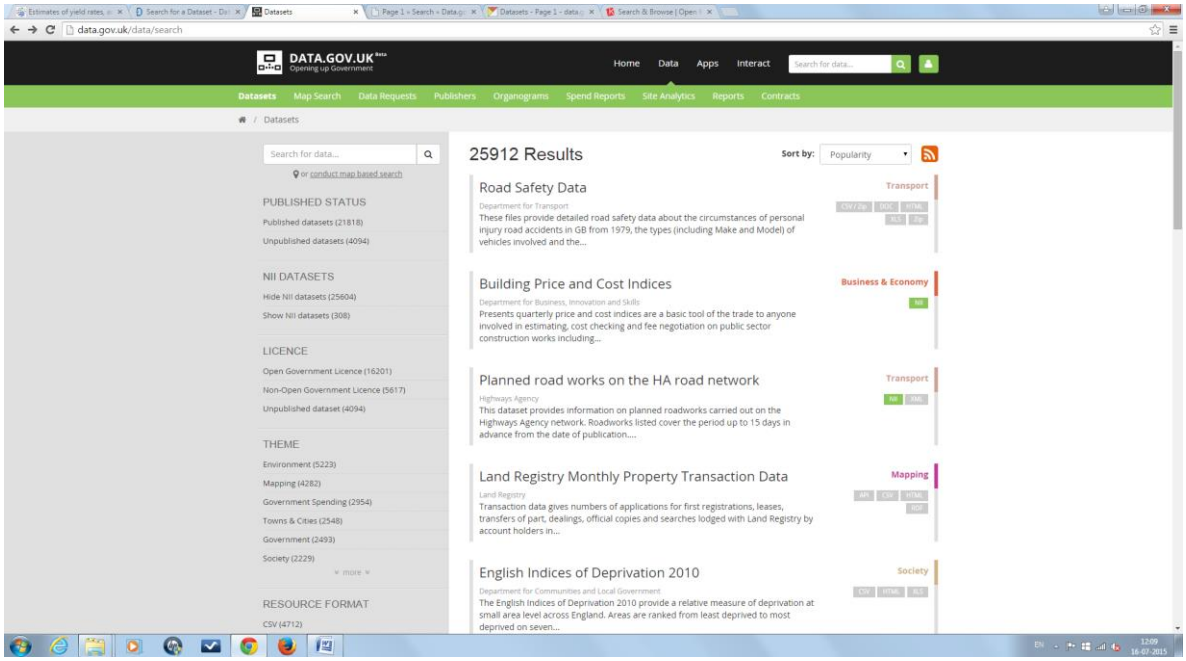


Figure 6: Screenshot of data.gov.uk

2.6 Opendata.go.ke (Kenya)[13]

Kenya is the first developing country to have an open government data portal. After Morocco it is the second on the continent and first in sub-Saharan Africa. The initiative has been noted globally as one of the most important steps Kenya has made to advance governance and also the implementing the new Constitution's provisions on information access.

Till November 2011, the approximate number of datasets uploaded to the site is close to 390 with a plan to upload more data over the upcoming year. The approximate number of page views is over 17,000 and over 2,500 dataset downloaded and embedded to various websites and portals.

Kenya's information is a national asset, and this site is about sharing it. The goal of opendata.go.ke is to make core government demographic, expenditure, development and statistical data available in a digital format which would be useful for researchers, ICT developers, policymakers and general public.

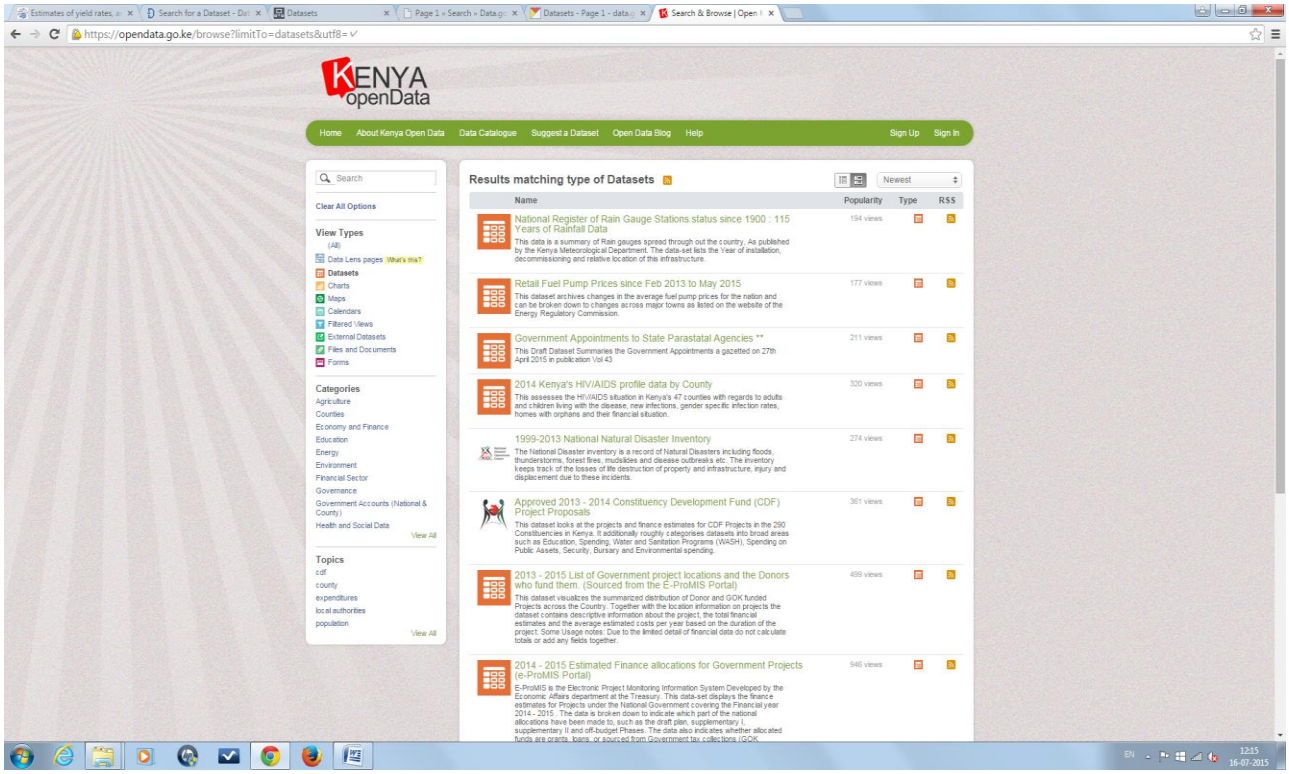


Figure 7: Screenshot of opendata.go.ke

3. Comparative study between different government sites

To make a comparative study between different data government sites, I have selected few categories like:

- **Topics/Sector/Categories:** An area/topics/sectors covered by a particular government.
- **Formats:** A file format is a standard way that information is encoded for storage in different data government sites.
- **Metadata for Dataset:** Metadata is defined as data about data. In a dataset metadata is useful to understanding and interpreting the contents of the dataset.
- **Purpose:** What is the main aim for providing data for the public?
- **Data Category:** Under data category, I have selected few criteria like year of commencement, total number of dataset available, majority of dataset in a particular domain, technology used for publish the data, developer and maintainer of the dataset etc.

Here, I have shown the different categories in a tabular format.

3.1. Topics/Sector/Categories

data.gov.in	data.gov	data.gov.uk	data.gov.au	data.govt.nz	data.go.ke
1. Water resources 2.Agricultural marketing 3.Family Welfare Statistics 4.Education 5.Environment and Forest 6.Crime Statistics 7.Health 8.Health and Family Welfare 9.Health Management Information System	1.Agriculture 2.Business 3.Climate 4.Consumer 5.Eco System 6.Education 7.Energy 8.Finance 9.Health 10. Local	1.Environment 2.Mapping 3.Government Spending 4.Towns & Cities 5.Government 6.Society 7.Health 8.Education 9.Transport	1.Community Services 2.Business Support 3.Science 4.Environment 5.Sports and Recreation 6.Finance Management 7.Health care	1.Agriculture, forestry and fisheries 2.Arts, culture and heritage 3.Building, construction and housing 4.Commerce trade and	1.Agriculture 2.Counties 3.Economy and Finance 4.Education 5.Energy 6.Environment 7.Financial Sector 8.Governance 9.Government

10.Road Transport 11.Rajya Sabha & Lok Sabha English/Hindi RSS feed 12.Labour and Employment 13.Socio-Economic 14.Roads 15.Financial Resources 16.Road Accidents 17.Telecommunication 18. Elementary education 19.Economic Survey and SDP of MadhyaPradesh 2013-2014 20.Rajya sabha Special Mentions 21.Rural Roads 22.State Power utilities and Electricity Departments 23.Companies 24.Houselisting and Housing Census data 25.Power and Energy 26.Union Budget 2013-14,Expenditure budget 27.Indian railways 28.Prison Statistics 29.Higher Education Statistics 30.PNG economic and Statistics 31.Rural Health Statistics 32.Rural Development 33.Transport 34.Higher Education 35.School education Statistics 36.Power and Energy 37.Social Development 38.Agriculture 39.Energy 40.India HDR 41.S&T financial resources and human resources 42.Housing and Urban affairs 43.Members of Loksabha 44.DRDO product 45.Sanitation	Govt. 11.Manufacturing 12.Ocean 13.Public Safety 14.Science & Research	10.Business & Economy	8.Civil Infrastructure 9.Cultural Affairs 10.Communications 11.transport 12.Employment 13.Government 14.Education and Training 15.GovHack 16.Governance 17.Geography 18.Emergencies 19.Tourism 20.Society 21.Indigenous Affairs 22.Safety 23.Property 24.Primary Industries 25.News 26.Law 27.Technology 28.Planning 29.Natural Resources 30.Information Communication 31.Immigration 32.General	industry 5.Education 6.Employment 7.Energy 8.Environment and conservation 9. Fiscal, tax and economics 10.Health 11.Infrastructure 12.Justice 13.Land 14.Local and regional government 15.Māori and Pasifika 16.Migration 17.Population and society 18.Science and research 19.State sector performance 20.Tourism 21.Transport 22.Ministers, cabinet and portfolios	Accounts 10.Health and Social data 11.Infrastructure 12.Population 13.Water and Sanitation

Table1: Different domain categories available in data government sites

Remarks:

The various government data sites provide data on various topics. Many of them named as sectors, categories, groups, themes etc. But they are basically same. Like in case of data.gov.in (India), they provide data on forty five groups like health, education, roads, sanitation etc., followed by data.gov.au (Australia) thirty two groups, and data.govt.nz (New Zealand) twenty two categories, data.gov (USA) fourteen topics, data.gov.uk (UK) ten themes and data.go.ke (Kenya) thirteen categories. From this table it is clear that all the enlisted data government sites are providing similar kind of information's for their respective countries may be in different names.

3.2. Formats

data.gov.in	data.gov	data.gov.uk	data.gov.au	data.govt.nz	data.go.ke
1.CSV, 2.Text 3.Application/vnd.ms-excel 4.XML, 5.JSON, 6.JSONP, 7.XLS, 8.ODS, 9.HTML, 10.WMS, 11.Application/Zip, 12.Spreadsheet	1.CSV, 2.XML, 3.XLSX, 4.PDF, 5.HTML, 6.API, 7.TSV, 8.JSON, 9.ZIP, 10.GIF, 11.TXT, 12.KML, 13.KMZ, 14.RDF, 15.Application/x-troff, 16.Audio/basic, 17.Application/simple, 18.Application/vnd.lot, 19.Application/octed-s, 20.Application/jpg, 21.WMS, 22.Excel, 23.WFS, 24.NetCDF, 25.KML, 26.TXT, 27.GML, 28.Esri REST, 29.API, 30.Application/tif,	1.RDF, 2.ZIP, 3.ODS, 4.SPARQL 5.CSV, 6.XML, 7.XLS, 8.PDF, 9.HTML, 10.API, 11.TSV, 12.JSON, 13.GIF, 14.TXT, 15.WMS,	1.Application/zip, 2.SHP, 3.Audio/basic, 4.Plain, 5.Text/Html, 6.Application/Pdf, 7.KMZ, 8.CSV, 9.XML, 10.XLS, 11.PDF, 12.HTML, 13.API, 14.TSV, 15.JSON, 16.ZIP, 17.GIF, 18.TXT, 19.KML, 20.SHAPEFILE, 21.WFS, 22.WMS, 23.Arcgrid, 24.Doc, 25.XLSX, 26.Multiple 27.Spatial, 28.Xml, 29.GeoJSON, 30.TXT, 31.Application/vnd.ms, 32.Image/jpeg, 33.Metadata, 34.Application/vnd.open, 35.RSS,	1.XLS, 2.KML, 3.API, 4.OtherGeo, 5.XML, 6.DB, 7.CSV, 8.PDF, 9.HTML, 10.Spreadsheet	1.CSV, 2.RDF, 3.RSS, 4.XLSX, 5.XML, 6 PDF, 7 JSON, 8.XLS,

	31.WCS, 32.Esri shapefile, 33.Tiff, 34.Mrsid, 35.XLS, 36.Fema-dcs-hydrology, 37.Fema-dcs-hydraulics, 38.arce, 39.TAR, 40.Fema-dcs-terrain, Ascii, 41.Fema-dcs-survey, 42.Export, 43.XYZ, 44.Application /vnd.goo, 45.Geotiff, 46.Application /xslt+xml, 47.Tgrshp, 48.Esri geodatabase fe, 49.Binary, 50.Mr Sid, 51.SHAPEFILE		36..Csv, 37.Docx, 38.API		

Table 2: Different formats available in data government sites

Remarks:

This table shows that the datasets for data government sites are available in several formats except for the New-Zealand, India and Kenya. New Zealand provides their datasets in ten formats only, India providing their dataset in twelve formats only and Kenya provides eight formats only.

On the other hand, data.gov (USA) provides data on fifty one formats, data.gov.au provides in thirty eight formats and data.gov.uk (UK) in fifteen formats. So, it is clear from the table that data.gov has most number of formats.

3.2.1 Most Commonly used Formats

Sl No.	Data.gov Formats	data.gov.in	data.gov	data.gov.uk	data.gov.au	data.govt.nz	data.go.ke
1.	CSV	√	√	√	√	√	√
2.	XML	√	√	√	√	√	√
3.	JSON	√	√	√	√	x	√
4.	JSONP	√	x	x	x	x	x
5.	XLS	√	√	√	√	√	√

6.	ODS	√	x	√	x	x	x
7.	PDF	x	√	√	√	√	√
8.	HTML	√	√	√	√	√	x
9.	API	x	√	√	√	x	x
10.	TSV	x	√	√	√	x	x
11.	RDF	x	√	√	x	x	√
12.	ZIP	√	√	√	√	x	x
13.	GIF	x	√	√	√	x	x
14.	TXT	√	√	√	√	x	x
15.	KML	x	√	x	√	x	x
16.	KMZ	x	√	x	√	x	x
17.	SHAPEFILE	x	√	x	√	x	x
18.	WFS	x	√	x	√	x	x
19.	WMS	√	√	√	√	x	x
20.	Spreadsheet	√	x	x	x	√	x
21.	RSS	x	x	x	√	x	√
22.	API	x	√	x	√	x	x
23.	Image/Jpeg	x	√	x	√	x	x
24.	Tar	x	√	x	x	x	x
25.	Audio/Basic	x	√	x	√	x	x
26.	SPARQL	x	**x**	√	**x**	x	x
27.	Tiff	x	√	**x**	**x**	x	x

Table3: Most Commonly used formats by different data government sites

Remarks:

It is viewed that there are various formats available for datasets, out of which CSV, XML, JSON, XLS, HTML, PDF, ZIP, TXT, GIF and WMS formats are used by almost data government sites. There are other famous formats also used by some of the sites like RDF, SHAPEFILE, KMZ, etc.

3.3. Metadata for Dataset

data.gov.in	data.gov	data.gov.uk	data.gov.au	data.govt.nz	data.go.ke
1.Title 2. Date 3. File Size 4.Download 5.Frequency 6.Granularity 7. Download Url 8. Description 9. Keywords 10.Contributor 11.Sectors and Sub Sector 12.Published on data portals 13. Group name 14.Asset	1.Title 2.Resource Type 3. created data 4. updated date 5.Publisher 6.Unique Identifier 7. maintainer 8. Maintainer Contact 9.Public Access Level 10.Bureau Code 11.Metadata Context	1.Format 2. Resource 3.Quality Check 4.Url 5. Date Updated 6. Last Updated 7.Title 8.Added to data.gov.uk 9.Theme 10.Themes (secondary) 11.Mandate 12.Temporal coverage 13.Geographic	1.Field 2. Title 3. Type 4. Language 5.License 6. data status 7. landing page 8.Date Published 9.Date Updated 10.Contact Point 11.Temporal Coverage 12.Geospatial Coverage 13. Jurisdiction 14.Data Portal	1. Title 2.Dataset Url 3. Date list 4. Rights 5. Costs 6.Agency 7. Contact 8.Date created 9. Date updated 10.Frequency of Update 11. Category 12. Keywords 13.Email 14.Phone 15.File Identifier	1.Title 2. Permission 3.Tag 4.Url 5.Data Provider 6.Source 7.Contributor

Jurisdiction 15.Category 16.Access Method 17.Access type	12.Schema Version 13.Catalog Described by 14.Data Quality 15.Data Dictionary 16.Harvest Object Id 17.Harvest Source Id 18.Harvest Source Title 19.Languageen- 20.Date Last Update 21.Program 22.Source Data Json Identifier 23.Source Hash 24.Source Schema Version 25.Spatial	coverage 14.Schema/Vocabulary 15.Update frequency 16.Temporal granularity 17.Code list 18.Service Level	15.Publisher/Agency 16.Created 17.Format 18.Size 19.Data dict 20.Datastore active 21. Has view 22.Hash 23. Id 24.Last modified 25.mime type 26.On same domain 27.Package id 28.resource type 29.Revision id 30.Size 31.State 32.Url type 33.Webstore last updated 34.webstore url	16.Language 17.Character Set 18.Hierarchy level 19.Hierarchy Level Name 20.Date Stamp 21. Metadata Standard Name 22.Metadata Standard Version 23.Identification Info 24.Data Quality info 25.Metadata Constraints	

Table 4: Metadata for Dataset available in data government sites

Remarks:

It is clear from the table that the metadata are not same for all the data government sites. It varies from government to government. It is expected that the preferences are different, for different data government sites. It is provided according to the need of different data government sites. But it is one of the main problems to improve or achieve interoperability among metadata schemas for the purposes of facilitating conversion and exchange of metadata**.**

3.3.1 Most commonly used Metadata

Sl. No.	Data.gov Metadata	data.gov.in	data.gov	data.gov.uk	data.gov.au	data.govt.nz	data.go.ke
1.	Field	x	x	x	√	x	x
2.	Title	√	√	√	√	√	√
3.	Type	x	√	x	√	x	x
4.	Language	x	√	x	√	x	x
5.	License	x	x	x	√	√	x
6.	Date created	√	√	x	√	√	x
7.	Date Published	√	x	x	√	x	x
8.	Date Updated	x	√	√	√	√	x
9.	Contact Point	x	x	x	√	√	x
10.	Temporal Coverage	x	x	√	√	x	x

11.	Geospatial Coverage	x	√	√	√	x	x
12.	Jurisdiction	x	x	x	√	x	x
13.	Data Portal	√	x	x	√	x	x
14.	Publisher/Agency	√	√	x	√	x	√
15.	Url	√	√	√	√	√	√
16.	Keywords	√	x	x	x	√	√
17.	Schema	x	√	√	x	x	x
18.	File size	√	x	x	√	x	x
19.	Resource	x	√	√	√	x	x
20.	Data quality	x	√	x	x	√	x

Table 5: Most Commonly used Metadata in data government sites

Remarks:
This table shows that among all the metadata elements title, url, date updated, date created, type, resource, geo special coverage, publisher, is the most used metadata elements by different data government sites. There are also metadata like keywords; temporal coverage, language etc. are also used.

3.4 Purpose

1.	data.gov.in	The portal is intends to increase transparency in the functioning of Government and also open avenues for many more innovative uses of Government Data to give different perspective.
2.	data.gov	One can find Federal, state and local data, tools, and resources to conduct research, build apps, design data visualizations, and more.
3.	data.gov.uk	The Government is releasing public data to help people understand how government works and how policies are made.
4.	data.gov.au	The main purpose of the site is to encourage public access to and reuse of government data by providing it in useful formats under open licences.
5.	data.govt.nz	It provides an easy way to find access and reuse public datasets from Government.
6.	data.go.ke	The goal of site is to make core government development, demographic, statistical and expenditure data available in a useful digital format for researchers, policymakers, ICT developers and the general public.

Table 6: Purposes of data government sites

Remarks:
This table shows the purpose of different data government sites. It is more or less same for all the data government sites. Their main goal is to releasing public data to inform the citizens of the country that how government works and how policies are made for the benefit of the society.

3.5 Data Category

Sl. No	Categories	data.gov.in	data.gov	data.gov.uk	data.gov.au	data.govt.nz	data.go.ke
1.	Year of	2012	2009	2010	2013	2009	2011

	commencement						
2.	Number of dataset	15468	131196	19343	5951	3435	654
3.	Majority of the dataset	Water resources (566)	Commerce (38,854)	Environment (5216)	Geosciences (2843)	Geo (2440)	Counties (113)
4.	Metadata Standard	Unknown	Project open data schema	Unknown	Unknown	ANZLIC Metadata Profile: An Australian/New Zealand Profile of AS/NZS ISO 19115:2005, Geographic information - Metadata	Unknown
5.	Technology	The site is based on Drupal Framework	It is powered by two open source applications, CKAN and Wordpress	It runs on a mix of Drupal and Comprehensive Knowledge Archive Network (CKAN)	Comprehensive Knowledge Archive Network (CKAN)	Unknown	Socrata offers a free account to upload datasets in CSV format. Users can download a dataset or access data via API.
6.	Developer	NIC,Dept. of Electronics & IT, Government of India & Office of Citizen Services & Innovative Technologies, General Services Administration, U.S. Government	U.S. General Services Administration, Office of Citizen Services and Innovative Technologies	The Transparency and Open Data team in the Cabinet Office,UK	The Office of the Australian Government CTO in the Department of Finance	The Department of Internal Affairs,NZ	Powered by Socrata

Table 7: Different data categories for comparison of data government sites

Remarks:

This table shows the different data categories for the data government sites. It covers various elements e.g. year of commencements, number of datasets in different data government sites, metadata standards, technology used, etc. It could be seen that majority of them used Comprehensive Knowledge Archive Network (CKAN), the open source data portal software for data management system.
Also some of them uses Content Management Software (CMS) like Drupal, Wordpress etc. These are all mainly open source software. Only data.go.ke (Kenya) uses Socrata, which is proprietary software. Most of the sites were developed by the respective departments of the particular countries like for India it is a joint venture of NIC and Government of US, for New Zealand it is developed by Department of Internal Affairs. There are no metadata standards for most of the data government sites only few of them follow some standards.
It is seen that the majority of the dataset available for different data government sites are different. Most of them have different priority of the datasets like in US they provide most of the datasets about commerce; in UK most of the datasets are on environments.

4. Methodology and Analysis of the results

As, I have mentioned earlier that I have followed two methodologies- a) overall assessment of the status of OGD portals of different national governments and b) review of six selected cases of OGD portals.

4.1 Overall assessment of the status of OGD portals of different national governments:

The first approach is reviewing the status and progress of the national government's OGD portal worldwide. The review of these portals focused on the data representation, availability of metadata and availability of different formats etc.

This methodology has provided some results those are enlisted below:

i) The majority of OGD portals complied with the open data principles in terms of providing granular data, accessibility, and share ability of data.

ii) The OGD portals offer data in machine readable format thus increasing the likelihood users can share and manipulate data.

iii) A greater part of the OGD portals adopt an Open License agreement that could increase the accessibility of data.

iv) Some of the portals hosted online or mobile applications in their portal.

v) Few of the OGD portals provide features facilitating the users to send request for dataset suggestions.

vi) The majority of the portals commonly use social media to support their reach and engagement effort.

vii) Few of the portals provide visualization features, from limited basic charts and maps to advanced charting and visualization.

4.2 Review of six selected cases of OGD portals:

The selection of the six portals also followed convenience sampling logic (It is a non-probability sampling technique where samples are selected because of their convenient accessibility and proximity) by primarily selected countries in the continent of world. The selected countries and multinational organizations are: India (Asia), United States of America (America), United Kingdom (Great Britain), Australia, New Zealand, Kenya (Africa).

The second methodology has provided some interesting results those are as follows:

i) It is seen from the table that the oldest data government sites were data.gov and data.govt.nz (2009). The newest one is data.gov.au, which was started in the year 2013. It is also seen that day by day most of the countries are publishing their data for public like data.gov.uk (2010), data.go.ke (2011), data.gov.in (2012).

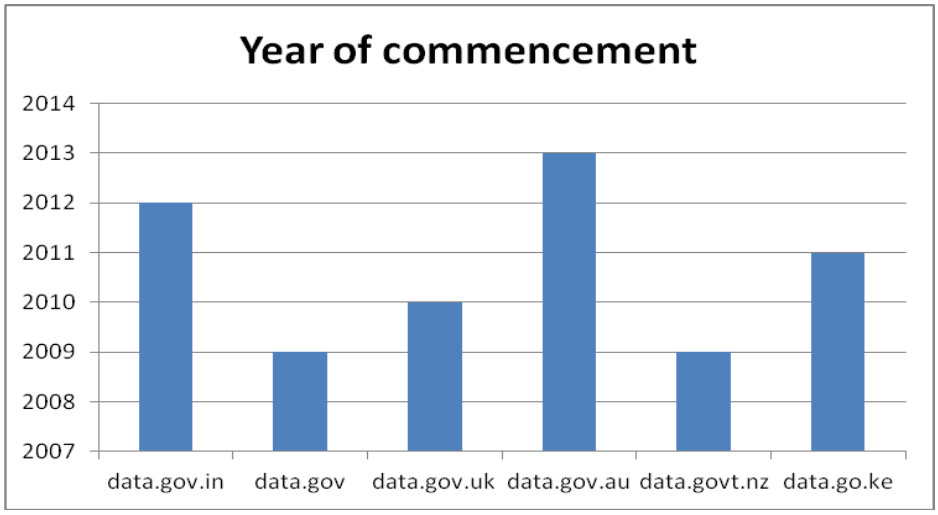


Figure 8: Year of commencement of government data sites

ii) Data.gov (USA) provides most number of datasets (131196) among all the listed countries. The second largest dataset provider is data.gov.uk (19343). India is ranked third among them, which provides 15468 datasets.

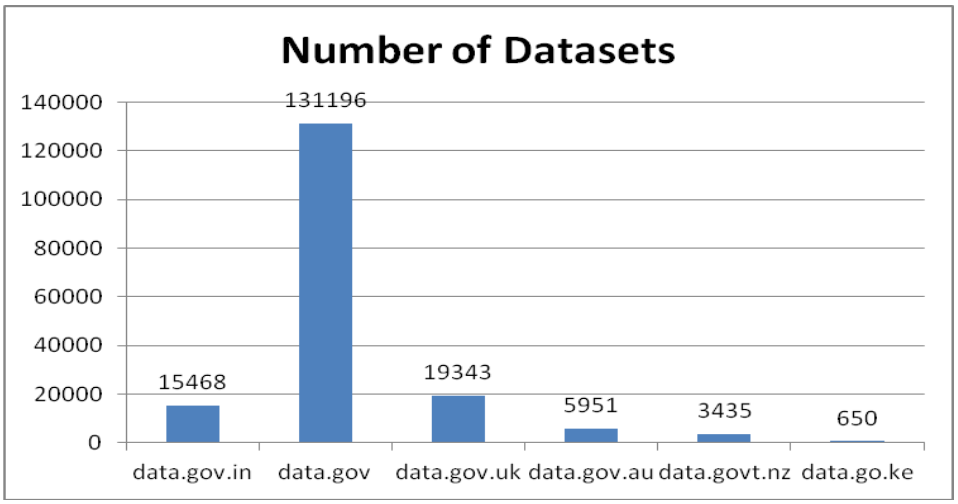


Figure 9: Number of datasets of different data government sites

iii) It is seen from the table that different data government sites have different set of majority datasets. Data.gov(USA) has majority of dataset on commerce(38854), data.gov.in has majority of datasets on water resources (566), data.gov.uk has 5216 datasets on environment, data.gov.au has 2843 datasets on geo science,data.govt.nz (Geo,2440), data.go.ke (counties,113) respectively.) It is visible that some of the data government sites focus more on a particular sector.

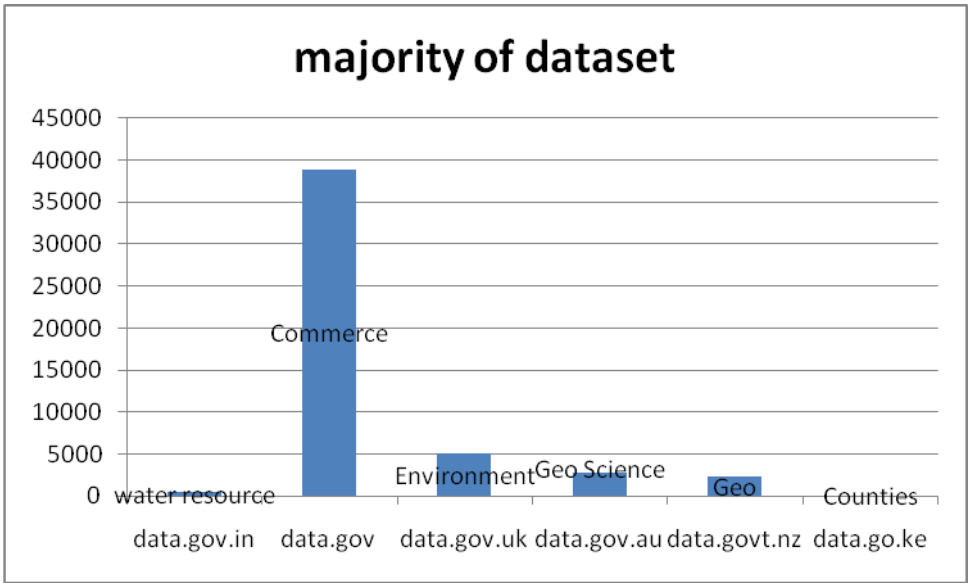


Figure 10: Majority of the Dataset

iv) Among the six data government sites data.gov (USA) provides maximum number of data formats followed by data.gov.au (Australia). Most of the data government sites provide data in different formats, among them CSV, XML, HTML, JSON are the most common formats used by majority of the data government sites. There are also some uncommon formats for different data government sites like data.gov.in (India) provides ODS format, data.gov (USA) provides SHAPEFILE, KML format, data.gov.uk (UK) provides WMS, GIF format, data.gov.au (Australia) provides WFS, KMZ formats etc.

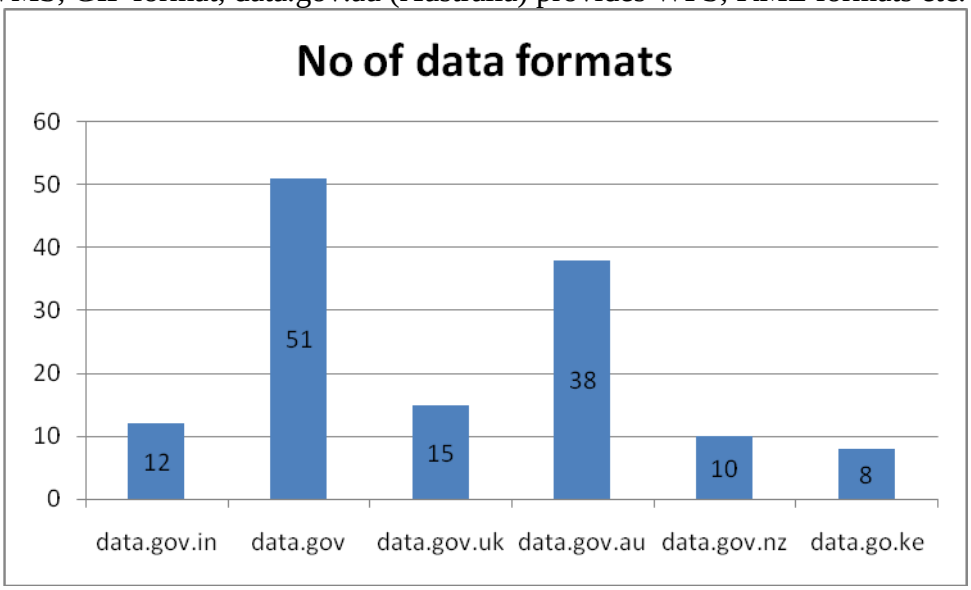


Figure 11: Number of data formats in different data government sites

v) It is seen from the table that data.gov (USA) provides most of the metadata for their datasets, almost fifty one metadata, followed by data.govt.au (Australia) thirty eight and data.gov.uk (UK) fifteen metadata elements. As the metadata provided by different data government sites are different so interoperability among the data sites are difficult.

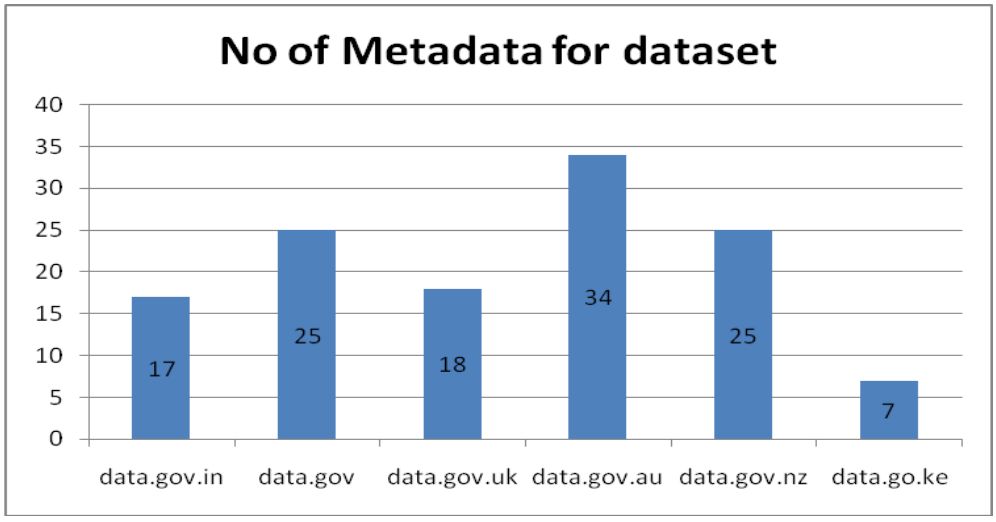


Figure 12: Number of Metadata for datasets in different data government sites

vi) It is clear from the table that data.gov.in (India) covered forty five topics, whereas data.gov.au (Australia) covered thirty two topics, data.govt.nz (New Zealand) covered twenty two topics respectively. The broad division is different for different data government sites. They have their own way of categorising the topics but the area covered may be same, although it was mentioned earlier that their focus is different.

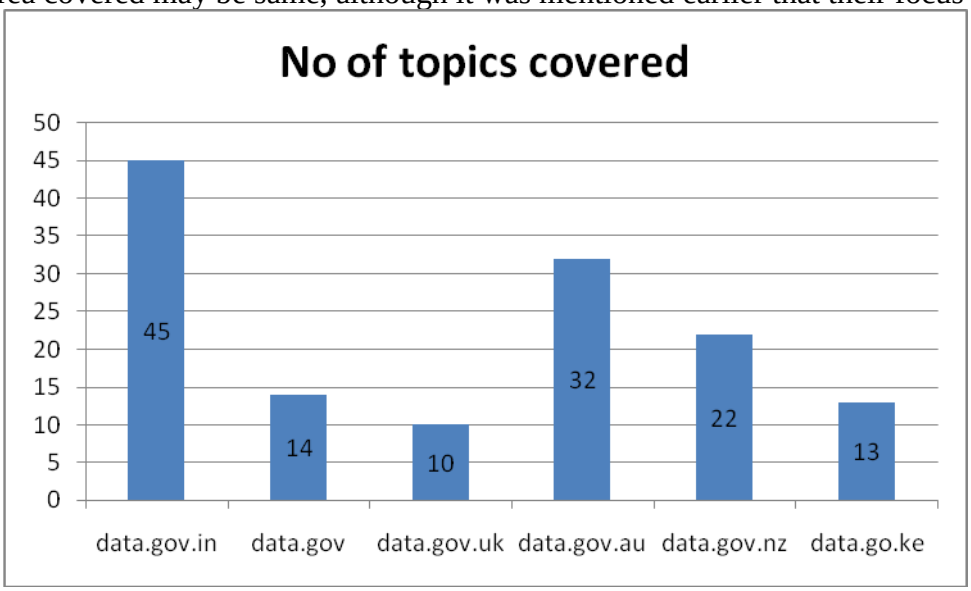


Figure 13: Number of topics covered by different data government sites

vii) It is very interesting that most of the data government sites use open source software for their data portals. It seems that they are also supporting open source rather than proprietary software.

viii) The metadata standards are different for different data government sites. Also, some of them are not following any metadata standards like in data.gov.in, data.go.ke, there is no metadata standard.

5. Framework for model data.gov.in site

I would like to propose a framework which was first suggested by Dr. Devika P. Madalli and Mr. Sudipta Biswas in Sci-data conference, 2014, New Delhi. The framework can be divided in two parts: Data Provider and Service Provider.

5.1 Data Provider:

i) The first part is for data providers to host their data in a standard way so that in future others can use/reuse the data more semantically, where data provider is hosting the structured datasets formats like CSV, XML, JSON etc.

ii) Once the data is made available, it is required to be described using a metadata schema (e.g. Dublin Core) or a vocabulary for metadata (e.g. DCAT).

iii) Now to expose the metadata to any harvester service or a service provider, the site should be complaint to OAI-PMH (Open Access Initiative- Protocol for Metadata Harvesting) and OAI-ORE (Open Access Initiative- Object Reuse and Exchange). Both protocols are http complaint and complement each other. OAI-PMH allows gathering metadata and OAI-ORE allows gathering actual data.

5.2 Service Provider:

The second part is for service providers.

i) The service providers can collect both data and metadata so that they can offer web services using the information available with metadata and data.

ii) From metadata one can identify open linked data by categorising the datasets and exploring the relations between the data sets. One can even convert the data into RDF format.

iii) By harvesting the actual data in RDF format using OAI-ORE, it possible to merge the similar data sets and SPARQL queries to generate specific answers to queries.

iv) As data in RDF format is captured from various data repositories, it should not be difficult to develop Mashup application or Intelligent Agent Applications to provide data services.

v) Once the data, metadata and protocols are in place one can register site in a registry/directory of data repositories.

6. Conclusions

The framework is not completely new; it is only an adaptation of model that is being used in harvesting metadata and digital objects. Data.gov.in should follow the Project Open Data schema (includes required fields like Title, Description, Tags, Update, Publisher, Contact Name, etc.) for every data set displayed on government data sites. This schema uses DCAT, which is a RDF vocabulary designed to facilitate interoperability between data catalogs published on the web. DCAT is used to describe datasets in data catalogs, publishers increase discoverability and enable applications easily to absorb metadata from multiple catalogs. It further enables decentralized publishing of catalogs and drives federated dataset search across sites[4].After analysing the metadata used by different data government sites, I would suggest below enlisted metadata elements should include in different datasets available data government sites:

Sl. No	Metadata elements	Descriptions
1.	Title	A name given to the dataset
2.	Identifier	A unique identifier of the dataset
3.	Depositor	A person a who deposit the data
4.	Principal investigators	Lead researchers for a particular well-defined dataset
5.	Sponsors	A person or organization that pays for or contributes to the costs involved
6.	Subject	The topic of the content of the resource
7.	Rights	Information about rights held in and over the distribution
8.	Keywords	A keyword or tag describing the dataset
9.	Type of resources	The nature or genre of the content of the resource
10.	Publisher	An entity responsible for making the dataset available
11.	Method of data collection	It is the process of gathering and measuring data
12.	Sources	A reference to a resource from which the present resource is derived
13.	Units	An individual component of a larger or more complex dataset
14.	Format	The file format of the distribution
15.	Language	The language of the dataset
16.	Date of Publication	A date associated with the publication of the data
17.	Date of Modification	Most recent date on which the distribution was changed, updated or

		modified
18.	Country	The dataset published from which country
19.	Geography	The dataset published from which place of a country
20.	Kind of data	It is a classification identifying one of various types of data, such as audio, video, alphanumeric etc.
21.	Abstract	A short summary of the dataset
22.	License	This links to the license document under which the distribution is made available
23.	Column names	The names of the columns associated with the tables of the dataset
24.	Contact person	A person who is designated for giving information or being a representative for an organization
25.	Maintainer	The responsible authority for maintenance of dataset
26.	Byte Size	The size of a distribution in bytes
27.	Time period	The period covered by the dataset
28.	Datasets	A collection of data, published or curate by a single agent, and available for access or download in one or more formats
29.	Download url	A file that contains the distribution of the dataset in a given format
30.	Catalog	A data catalog is a curate collection of metadata about datasets

Table 6: List of suggested metadata elements and their descriptions

6.1 Justifications

Data.gov.in provides only few metadata for their dataset (e.g. Title, Date, File Size, Download, Frequency, Granularity, and Download Url). So it is expected that only these metadata are not sufficient for a dataset. We have to consider more number of metadata elements. For example, to describe a dataset we need to use "dataset" metadata element which describe a collection of data, published or curate by a single agent, and available for access or download in one or more formats. The "catalog" is also a much required metadata elements for data government sites which describe a curate collection of metadata about datasets. "Subject" is also a very important metadata element; we need to consider which describes the topic of the content of the resource. Like that all the enlisted metadata elements has their importance and are essential to describe a resource.

6.2 Suggestions[18]

- Attention should be given to the correlation between focus of data being published and the public need.
- Governments need to invest significant time to allow agencies to prepare data for publication.
- Government need to consider which format to publish their data. There are many elements to consider the format of data to be published, such as: the format of currently available data, availability of resources and technical capability and others.

The types of user engagement provided in the OGD portal might correlate to the level of user's engagement. The review indicates different types of user's engagement and participation provided in the OGD portal, from user ratings to community-based engagement.

6.3 Limitations of the study

There are some limitations of this study. These are

- Only six data government sites have been considered.
- Convenience sampling logic (non probabilistic sampling) is followed for selecting the samples, which is not a good way of sampling, biasness may occur.
- Only few of the categories were considered for comparative study.
- The selection of government data sites could be better.

- Working procedure of data provider and service provider is not mentioned.
- A practical implementation of the proposed framework is missing.

In spite of this kind of limitations, I hope that the proposed framework may work in the future. It is expected that some government sites may follow the framework and the metadata (mentioned earlier) in the near future.

7. Acknowledgements

I wish to express my deep and sincere gratitude to my guide Dr. A.R.D Prasad, Professor, Documentation Research and Training Centre (DRTC), Indian Statistical Institute, Bangalore Centre (ISIBC) for his inspiring guidance and overall supervision. His constant encouragement and ideas that he generously shared with me helped immeasurable at every stage of this dissertation. I am also expressing my thanks to Dr. Devika P. Madalli for her valuable suggestions regarding selection of this topic. I would like to offer my thanks to Mr. Sudipta Biswas, SRF, DRTC for his cooperation and encouragement in completing this paper.

8. References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In *The semantic web* (pp. 722-735). Springer, Berlin, Heidelberg.
2. Braunschweig, K. et al (2012). The State of Open Data: Limits of Current Open Data Platforms. Retrieved from http://www2012.org/proceedings/nocompanion/wwwwebsci2012_braunschweig.pdf
3. Chan, L. M., & Zeng, M. L. (2006). Metadata interoperability and standardization–a study of methodology part I. *D-Lib Magazine*, *12*(6), 1082-9873. Retrieved from http://www.dlib.org/dlib/june06/chan/06chan.html.
4. Data Catalog Vocabulary. (2018). Retrieved January 10, 2018 from http://www.w3.org/TR/vocab-dcat/
5. Data.gov. (2018). Retrieved January 18, 2018 from http://data.gov/.
6. Data.gov.au. (2018). Retrieved December 23, 2017 from http://data.gov.au/.
7. Data.gov.in (2018). Retrieved January 22, 2018 from http://data.gov.in/.
8. Data.gov.uk. (2018) Retrieved February 15, 2018 from http://data.gov.uk/.
9. Data.govt.nz. (2018). Retrieved January 31, 2018 from http://data.govt.nz/.
10. Davies, T. (2010). Open Data, Democracy and Public Sector Reform: A look at Open Government Data use at data.gov.uk. Retrieved January 12, 2018 from www.opendataimpacts.net/report/wp-content/uploads/2010/08/How-isopen-government-data-being-used-in-practice.pdf.
11. Dekkers, M., Polman, F., te Velde, R., & de Vries, M. (2006). Measuring European Public Sector Information Resources. *Final Report of Study on Exploitation of public sector information–benchmarking of EU framework conditions*. Retrieved December 15, 2018 from https://ec.europa.eu/digital-single-market/en/news/mepsir-measuring-european-public-sector-information-resources-final-report-study-exploitation-0.
12. Free File Format Definition. (2007). Retrieved February 15, 2018 form http://www.linfo.org/free_file_format.html.
13. Kenya Open Data. (2018). Retrieved February 05, 2018 from https://opendata.go.ke/.
14. Mustafa, M. (2012). Corporate Use of Open Government Data. Retrieved December 13, 2018 from https://www.open3.at/wp-content/uploads/2012/06/MM-OGD-JH-Final-v2.pdf.
15. Nugroho, R. (2013). A comparison of open data policies in different countries: Lessons learned for an open data policy in Indonesia. Retrieved from http://repository.tudelft.nl/view/ir/uuid%3Aae4e0a64-579d-40c4-bed0-d51614ddea9c/.
16. Open Government Data Principles. (2015). Retrieved December 13, 2017 from

https://public.resource.org/8_principles.html.

17. Open Government Data. (2015). Retrieved December 03, 2017 from http://opengovernmentdata.org/about/.
18. Sayogo, D. S., Pardo, T. A., & Cook, M. (2014, January). A framework for benchmarking open government data efforts. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 1896-1905). IEEE.
19. Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance*, (22), 0_1.
20. Worldwide Open Data Sites. (2013). Retrieved December 13, 2017 from http://dataremixed.com/2013/08/worldwide-open-data-sites/.