Open Data Manual

OPEN
KNOWLEDGE
NEPAL

About Open Knowledge Nepal

Open Knowledge Nepal is an open network of open knowledge enthusiasts. We are a non-profit organization comprised of openness aficionados, mainly self-motivated youths who believe that openness of data is powerful in order to have a participatory government with civil society, eventually leading to sustainable development. Those self-motivated people are social activists, professors, students, journalists, governments officers and more. The group has been involved in research, advocacy, training, organizing meetups and hackathons, and developing tools related to Open Data, Open Government Data, Open Source, Open Education, Open Access, Open Development, Open Research and others. The group also helps and supports Open Data entrepreneurs and startups to solve different kinds of data related problems they are facing through counseling, training and by developing tools for them.
Website: http://oknp.org

Table of Contents

1. Introduction

The increasing number of civil society and private organizations working for and with open data has made the momentum of open data stronger in Nepal. But despite all the positive developments, the global open data survey like the Global Open Data Index [2] and the Open Data Barometer [3] suggest that Nepal still has an evident need for human resources, with increased awareness and capacity to produce and use more open data for Nepal's progress in development, governance and other fields. The Government of Nepal has been generating and sharing massive amounts of data, but most of the released data sits in closed formats like PDF with other dependencies, which need to be processed and analyzed before being able to make proper use of it.

To help improve the data literacy of the human capital in Nepal, this Open Data Manual is compiled by Open Knowledge Nepal as a reference and recommended guide for university students, civil society and private sector. The manual is a compilation of information from a series of online resources covering key data topics to explain open data - its benefits and technicalities - to those who are interested to use open data to develop innovative tools and add value to products and services, as well as to those who are planning to join the open data ecosystem through new startups and ideas. The recommended readings, references and the guidance provided by the manual can be a great pool of resources for those interested in open data to improve their capacity, adding value in governance, education, civic engagement, journalism and more.

The Open Data Manual is especially focused on Nepal's digital natives (young generation), who want to know more about the background, history, sources, stories and process to extract, analyze, clean and visualize the available data in Nepal, making it distinct from the Open Data Handbook. The manual will make the Nepali youth more aware of the benefits of open data to them, will fill in the gap of data education, and will better prepare young people for a rapidly changing data scenario. For them who want to learn more about the what and why of open data, but do not want to jump on its technical side, we recommend going through the Open Data Handbook - "*a blueprint developed by Open Knowledge International, which has been used by government and civil society organization of different countries around the world*" [4].

Open Knowledge Nepal designed this manual as part of a broader project of training and workshops aimed at raising open data awareness to Nepal's digital natives. This project is supported by the Data for Development Programme, implemented by the Asia Foundation in partnership with Development Initiatives, with funding from the UK Department for International Development to improve the sharing and use of data as evidence for development in Nepal.

[2] Open Knowledge Internation (2017). *Global Open Data Index*. [online] Available at: https://index.okfn.org/ [Accessed 8 Oct. 2017].

[3] World Wide Web Foundation (2017). *Open Data Barometer.* [online] Available at: http://opendatabarometer.org/ [Accessed 8 Oct. 2017].

[4] Open Knowledge International (n.d). *The Open Data Handbook*. [online] Available at: http://opendatahandbook.org/ [Accessed 3 Oct. 2017].

2. Introduction to Open Data

What comes to your mind when you hear the term "open data"? A simple thought might be: data that is open. But "open" is a vague term in itself. How would you define open? Why do you need data to be open anyway? Imagine: on your way home, you see a pothole in the road. You remember the road was not built long ago, and it should not have a pothole. Then you begin to wonder who the contractor was, and how much of taxpayers' money was invested in building the road. Now, think if you really wanted to know who built the road that leads to your home, and how much was spent building the road, would you be able to do it easily? Who would you ask? Where would you search for the information?

Every fiscal year, the Government of Nepal announces its budget - information that is crucial to the national economy and livelihoods for the public. If the announcement is only done on TV or radio, the ways in which the released information can be used is limited. If the budget data was also released as spreadsheets, it would be possible for economists or even people with basic data literacy skills to understand and perform operations on the data to generate valuable insights. If data can be accessed by everyone, and in a format that enables reuse, it can have more value. If data is open it can be used in many different ways, like in the creation of smartphone applications, interesting visualizations and news stories. The co-founder of Open Knowledge International, Dr. Rufus Pollock believes, "*The best thing to do with your data will be thought of by someone else*" [5]. When data is open, the value can be added and created in ways previously unimagined.

Our life is guided by decisions we make and we need information to make decisions. The information we need can be about anything and everything. For example, you may want to know what emergency services are available in the nearest hospital, what options there are to study the degree of your choice, what the standard market price is of the goods you need to buy and more. However, the information you want might not be available, even though, in many cases, you have the right to that information as a citizen.

Data can be defined as information in its raw, pre-analyzed form, such as numbers, words, pictures etc. "**Open data refers to data that is made available in a machine readable format and shared publicly so that it is free to access, use and reuse for any purpose**" [6].

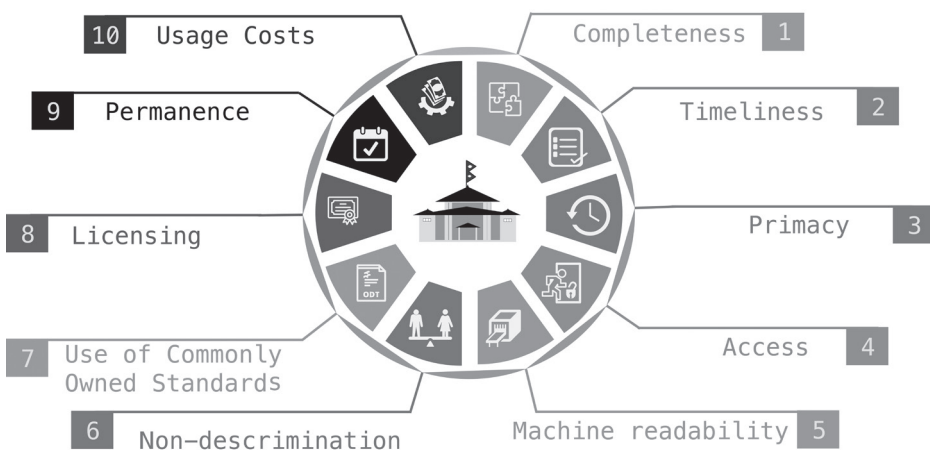[5] Pollock, R. (2017). *Miscellaneous· Rufus Pollock Online*. [online] Available at: https://rufuspollock.com/misc/ [Accessed 21 Sep. 2017].

[6] Open Knowledge International (n.d). *Open Definition.* [online] Available at: http://opendefinition.org/ [Accessed 22 Sep. 2017].

To explain it simply, with open data:

- The data itself should be accessible freely through the medium of internet such as through websites, data portals, etc.
- The available data should be usable and reusable without any legal restrictions.
- Using and performing operations on the data to add value through analysis, visualization, developing applications and more should be free.
- The data needs to be able to be shared freely and openly.

However, depending on the associated open data license, you might need to mention the original source of data and share the data and derivatives you make from the data in same or similar terms as the original data. The availability of data in an open and reusable format can add value to different groups of people, organizations, and government. It's untapped potential can be unleashed if public government data is turned into open data.

Principles of Open Data

10	Usage Costs	Completeness	1
9	Permanence	Timeliness	2
8	Licensing	Primacy	3
7	Use of Commonly Owned Standards	Access	4
6	Non-descrimination	Machine readability	5

Principles for opening up Government Information

The Sunlight Foundation lists ten helpful principles for assessing the extent to which government data is open and accessible to the public [7]:

---

[7] Sunlight Foundation (2010). *Ten Principles For Opening Up Government Information.* [online] Available at: https://sunlightfoundation.com/policy/documents/ten-open-data-principles/ [Accessed 19 Sep. 2017].

1. Completeness	2. Timeliness
3. Primacy	4. Ease of Physical and Electronic Access
5. Machine readability	6. Non-discrimination
7. Licensing	8. Use of Commonly Owned Standards
9. Permanence	10. Usage Costs

Slowly the Government of Nepal is also digitizing their works and had started sharing the collected information and data. This shows that government also wants to foster the new strategies to connect with citizens and conduct transparent day to day operations, and reap the benefits of open data.

Few benefits of open government data are
- Increases transparency and accountability of the government
- Develops trust, credibility, and reputation
- Promotes progress and innovation
- Encourages public education and citizens engagement
- Stores and preserves information

Not only government, but university students can also take direct and indirect benefits from open data

- It helps for research and new projects
- It supports new business growth
- It improves the skills of using new data tools and programming languages
- Students can use data and statistics for analysis and reporting
- Students can use data to build innovative solutions to tackle development challenges
- It gives access to wider range of learning from new sources

Recommended Readings

- Open Data Handbook (2017). *Why Open Data?.* [online] Available at: http://opendatahandbook.org/guide/en/why-open-data/
- Open Definition (n.d). *The Open Definition*. [online]Available at: http://opendefinition.org
- Sunlight Foundation (2010). *Ten Principles For Opening Up Government Information*. [online] Available at: https://sunlightfoundation.com/policy/documents/ten-open-data-principles/
- Open Data Institute (n.d). *What is open data?* [online] Available at: https://the-odi.org/what-is-open-data
- Srijana Acharya, Han Woo Park (2016). *Open data in Nepal: a webometric network analysis* [PDF] Available at Springer: https://link.springer.com/article/10.1007/s11135-016-0379-1
- World Wide Web Consortium (2009). *Publishing Open Government Data*. [online] Available at: https://www.w3.org/TR/gov-data/
- International Open Data Charter (2015). *Principles of open data.* [online] Available at: https://opendatacharter.net/principles/
- Opengovdata.io (n.d). *Open Government Data: The Book.* [online] Available at: https://opengovdata.io/
- Open Gov Partnership (n.d). *Resources | Open Government Partnership*. [online] Available at: https://www.opengovpartnership.org/resources

2.1. History and Background - Open Data in Nepal

You might have increasingly heard of terms such as "Big Data", "Citizen Generated Data", and "Open Data" over the last few years. The way data is conceived has changed in recent years, which can be credited to rapidly evolving technology and revelations of the power of data. For decision-making, especially by the government, good quality data in a useable format is required. In the past, enough data was not available to support evidence-based decisions, and decision-making largely depended on the experience of experts. Now, with evolving technology - information technology in particular - massive amounts of data are created daily.

The world we live in is a world rich in quantity, quality, and diversity of data, more than ever. The International Business Machines Corporation says 2.5 quintillion bytes of data are created every day [8]. This unravels a huge potential for innovation and value addition in decision-making and action for government, citizens, and businesses alike. In addition to innovation and decision-making, data is also necessary for tracking development progress. In 2015, nations of the world committed to achieve the Sustainable Development Goals (SDGs) by 2030. To track the progress made in reaching the targets of Sustainable Development agenda, a revolution in data production, sharing and use are required [9].

In Nepal, open data is still a relatively new concept and it has only been a few years since momentum for open data started growing. Over the last few years, there has been a growing number of new initiatives, organizations, and groups who have started working on addressing some of the challenges for open data in Nepal, and on improving the sharing and use of data in Nepal [10].
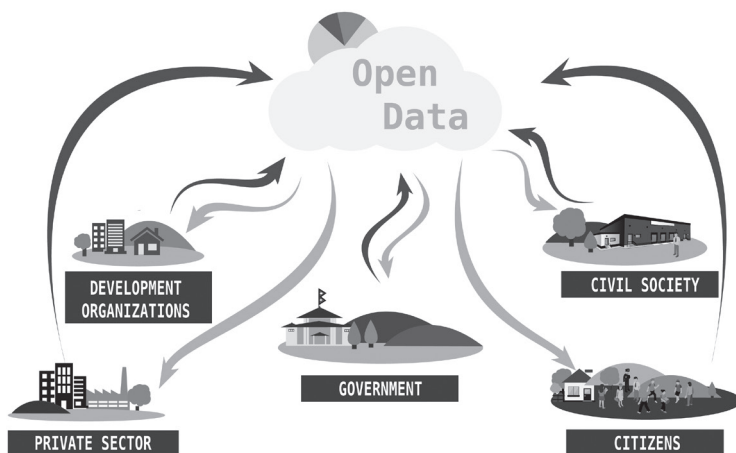
Nepal's open data ecosystem consists of actors from government, civil society, national/ international development organizations and the private sector. Different actors in government, civil society and private sectors have been engaging in notable scattered open data initiatives.

---

[8] Jacobson, R. (2013). *2.5 quintillion bytes of data created every day. How does CPG & Retail manage it?* [online] Available at: https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/ [Accessed 18 Sep. 2017].

[9] A World That Counts (2014). [PDF] *UN Data Revolution.* [online] Available at: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf [Accessed 2 Oct. 2017].

[10] Development Initiatives (2017). *Insights into Nepal's emerging data revolution.* [online] Available at: http://devinit.org/post/insights-into-nepals-emerging-data-revolution/ [Accessed 3 Oct. 2017].

Open Data Manual - Compiled by Open Knowledge Nepal

Nepal's Open Data Ecosystem

Open Data

DEVELOPMENT ORGANIZATIONS

CIVIL SOCIETY

GOVERNMENT

PRIVATE SECTOR

CITIZENS

Initiatives such as the large-scale mobile data collection to find out the damage caused by the earthquake [11], the National Planning Commission's self-assessment of data gaps [12] in measuring progress to achieve Sustainable Development Goals (SDGs), launch of Public Procurement Transparency Initiative in Nepal (PPTIN) by the Public Procurement Monitoring Office [13], sharing of registered business data by Office of Company Registrar [14] and the submission of the Open Government Data (OGD) National Action Plan to Government of Nepal by National Information Commission [15] clearly reflects Nepal government's presence in open data, mostly aided by the dynamic civic space. The participation of representatives from the Nepal Government in the Open Government Partnership summit in Mexico in 2015 was a sign of the government's interest in a mutual involvement in open data with civil society. The way forward is the use of open data for increased impact in national development and the involvement of a greater and diverse populace.

[11] Rohaidi, N. (2017). *Exclusive: How an earthquake led to Nepal's digital transformation* - GovInsider. [online] Available at: https://govinsider.asia/inclusive-gov/nama-budhathoki-kathmandu-living-labs-earth-quake-digital-transformation/ [Accessed 20 Sep. 2017].

[12] Nepal's Sustainable Development Goals (2017). [PDF] *Singha Durbar Kathmandu Nepal: National Planning Commission, Government of Nepal.* Available at: http://www.npc.gov.np/images/category/SDGs_Baseline_Report_final_29_June-1(1).pdf/ [Accessed 21 Sep. 2017].

[13] Open Nepal (2017). *Open Contracting: Paving the Way for Open Government in Nepal?* [online] Available at: http://opennepal.net/blog/open-contracting-paving-way-open-government-nepal [Accessed 2 Oct. 2017].

[14] OCR (n.d). *O.G.D - Office of Company Registrar.* [online] Available at: http://www.ocr.gov.np/index.php/en/data/o-g-d [Accessed 3 Oct. 2017].

[15] National Information Commission Official Site (2017). *Press Release 2074/05/07.* [online] Available at: http://www.nic.gov.np/page/press-release-20740507 [Accessed 20 Sep. 2017].

There have also been initiatives that have run in collaboration between government and civil society, leading to innovative mutual accountability practices. One of the great examples in Nepal's context is the Mapping and Opening Data for Local Governance and Citizen Engagement (MODEL4G) [16] project, initiated by Kathmandu Metropolitan City - Ward 7, Kathmandu Living Labs and Young Innovations, which uses mapping, open data, and mobile technology. Private sectors and development organizations are also playing a huge role in the open data ecosystem of Nepal. The competition like Data-Driven Farming Prize [17] and the Open Data Hackathon [18] are helping the growth of open data projects in Nepal. Civil society and private sectors initiative who are using data as a project backbone like DB2MAP [19], NepalMap [20], Election Nepal [21], Data Nepal [22], Nepal In Data [23], FACTS Nepal [24] etc., are getting right platform to showcase their works.

However, despite progress, challenges with regard to the production, sharing and use of high-quality data for su stainable development remains, and work needs to continue within all the different actors in the open data ecosystem to move the open data movement forward.

[16] Model4g (2017). *Ward 7 Kathmandu Metropolitan City*. [online] Available at: http://model4g.net [Accessed 21 Sep. 2017]

[17] Data Driven Farming Prize. http://datadrivenfarming.challenges.org/ [Accessed 2 Oct. 2017].

[18] Open Data Day Hackathon - Open Nepal. http://opennepal.net/odd2017/hackathon/ [Accessed 2 Oct. 2017].

[19] DB2MAP. http://www.db2map.com/ [Accessed 3 Oct. 2017].

[20] NepalMap: Explore and understand Nepal using data. https://www.nepalmap.org/ [Accessed 3 Oct. 2017].

[21] Election Nepal. http://electionnepal.org [Accessed 3 Oct. 2017].

[22] Data Nepal. http://datanepal.com [Accessed 3 Oct. 2017].

[23] Nepal in Data:: A Gateway to Development Data & Statistics in Nepal. https://nepalindata.com/ [Accessed 3 Oct. 2017].

[24] Facts Nepal. http://www.factsnepal.com/ [Accessed 3 Oct. 2017].

Recommended Readings

- A World That Counts (2014). [PDF] *UN Data Revolution.* [online] Available at: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf
- Dennison, L. (2014). *A data revolution in Nepal – starting the conversation.* [online] Development Initiatives. Available at: http://devinit.org/post/data-revolution-nepal-starting-conversation/
- Dennison, L. (2014). *Exploring the context for open data in Nepal.* [online] Available at: http://devinit.org/post/exploring-context-open-data-nepal/
- Dennison, L. and Rana, P. (2017). *Nepal's emerging data revolution.* [PDF] Development Initiatives. [online] Available at: http://devinit.org/wp-content/uploads/2017/04/Nepals-emerging-data-revolution.pdf
- Nepal's Sustainable Development Goals (2017). [PDF] *Singha Durbar Kathmandu Nepal: National Planning Commission, Government of Nepal.* [online] Available at: http://www.npc.gov.np/images/category/SDGs_Baseline_Report_final_29_June-1(1).pdf/
- Sapkota, K. (2014). *Exploring the emerging impacts of open aid data and budget data in Nepal.* [PDF] Freedom Forum. [online] Available at: http://www.opendataresearch.org/content/2014/724/exploring-emerging-impacts-open-aid-data-and-budget-data-nepal
- Central Bureau of Statistics (n.d). *Compendium of National Statistical System of Nepal.* [PDF] Available at: http://cbs.gov.np/image/data/2017/National%20Statistical%20System%20of%20Nepal_FINAL.pdf.
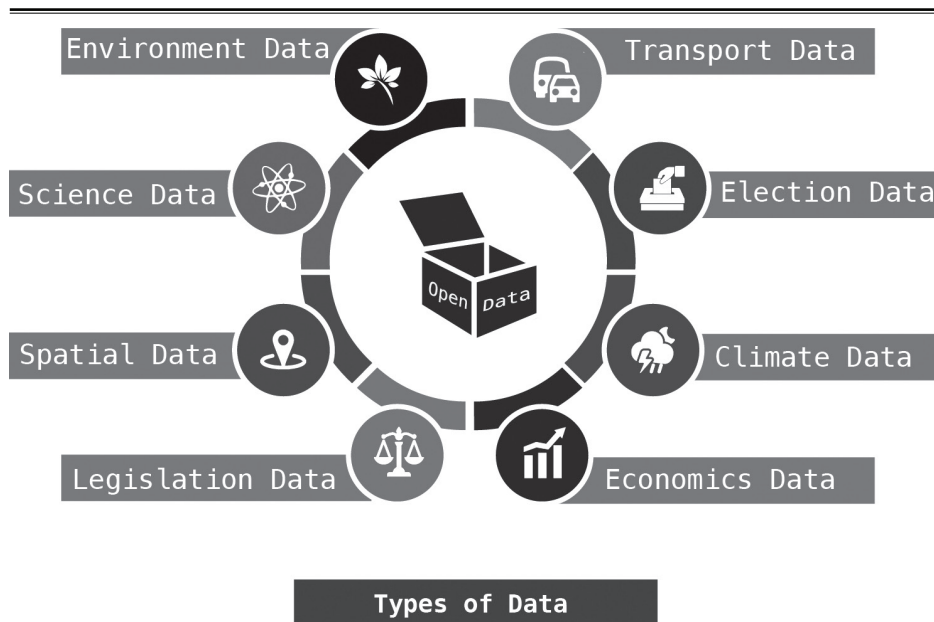
3. Working with Open Data in Nepal

Nepal currently does not have any official open data law and policy, but on August 25, 2017, the National Information Commission handed over a National Action Plan on Open Government Data to the Government of Nepal [25]. The submitted action plan was developed on the basis of discussion and interaction with RTI activists, government officials, civil society, statisticians, and academicians. The open government data action plan is currently under review.

Right to Information (RTI) Act 2007 [26] guarantees that Nepali citizens can access information on the functioning of any 'public body,' in order to make governance and policymaking more transparent and accountable. The National Information Commission (NIC) is responsible for the promotion and implementation of RTI in Nepal. Any Nepali citizen can submit a written RTI application to the Information Officer (IO) of the relevant public body. According to the Act, every public body must have a designated information office to respond to such requests and disseminate the associated information. If the application satisfies the requirements of the law, the information officer has to provide the information in the format requested by the applicant.

Since open data is still an emerging topic with only few key stakeholders, the great way to get started with open data in Nepal is to be connected with civil society organizations, who are working in Nepal to promote open data. Civil society organizations like Kathmandu Living Labs, Open Nepal, Open Knowledge Nepal, Freedom Forum, Code for Nepal, Accountability Lab, Bikas Udhyami etc., who work in policy research, advocacy, tech, journalism and many others fields relating to open data in Nepal. You can contact them for collaboration, consult, and volunteering opportunities.

---

[25] National Information Commission Official Site (2017). *Press Release 2074/05/07.* [online] Available at: http://www.nic.gov.np/page/press-release-20740507 [Accessed 20 Sep. 2017].

[26] Right to Information Act, 2065 (2009). [PDF] *Nepal Law Commission.* [online] Available at: https://www.moic.gov.np/acts-regulations/right-to-information-rules.pdf [Accessed 20 Sep. 2017].

Environment Data • Transport Data • Science Data • Election Data • Spatial Data • Climate Data • Legislation Data • Economics Data

Types of Data

In Nepal, there are different types of data, which are on demand and you can explore for opening up, such as Environment, Transport, Election, Climate, Economics, Legislation, Spatial, Science and so on. If you are working with data for your research, reports or just for analysis, we suggested you to follow the 8 simple steps to get started.

1. Brainstorm the types of data - which you want to open or explore.
2. Search for the datasets on the internet, through existing data portals, government websites or visit the respective government office physically.
3. Once you have desired datasets, explore it - if it's in PDF format, extract it and find the right pattern by giving it meta tags.
4. Clean the extracted data and analyze it - for this, you can use programming languages like Python and R to do analysis or simple spreadsheets formula to aggregate the rows and columns.
5. Make use of the data innovatively by integrating the data into your apps, research or simply by visualizing the data for sharing the outputs.
6. Use the data collected to tell a story - if you are good at writing, use visualization and articles for this purpose.
7. Share and publish the raw and analyzed data through websites, data portals, cloud storage so that others can also take benefits from it.
8. Make sure to provide an open license and attribution to your datasets, so that other users can reuse and add value to the created data without restrictions.
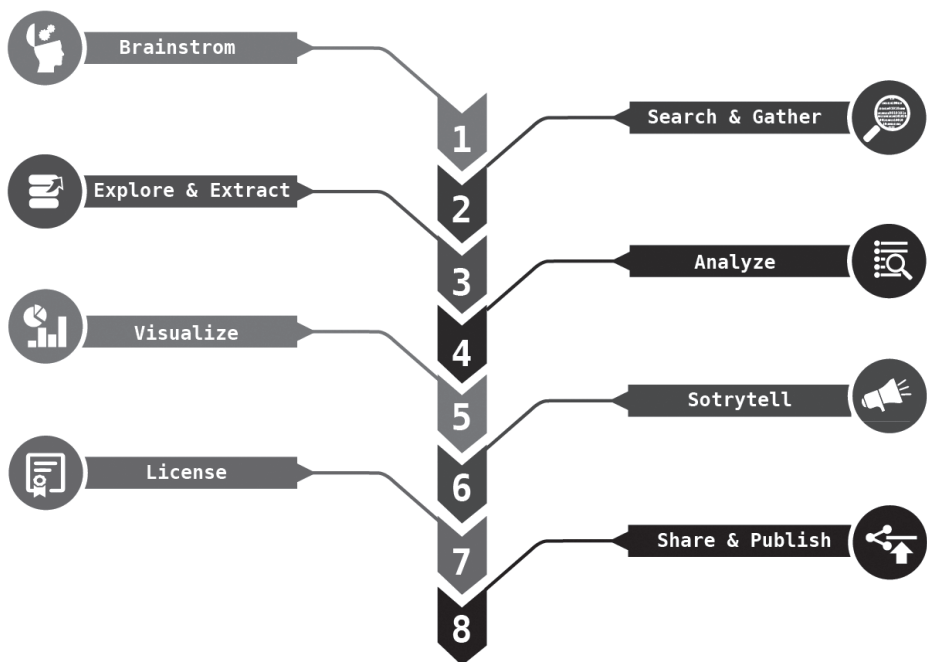
Working With Data

To learn more about the process of opening and using data, please refer to the technicalities section of this handbook that provides a detailed understanding on the 'how to' of the steps mentioned above.

While making use of the government data, please note that, Nepal currently does not have any kinds of licensing policy on data. Most of the data published by the Government of Nepal is "**All Rights Reserved**", which means users need to ask permission from the respective data provider before reusing it. Even with your work, taking credit or giving others a permission to use your work will all depend upon licensing. So if you are reusing datasets, if would better to ask permission with data provider once.

Recommended Readings

♦ Right to Information Rules, 2065 (2009). [PDF] *Nepal Law Commission.* [online] Available at: https://www.moic.gov.np/acts-regulations/right-to-information-rules.pdf

♦ Wiki.civiccommons.org (n.d). *Open Data Policy.* [online] Available at: http://wiki.civiccommons.org/Open_Data_Policy/

♦ Wiki.civiccommons.org (2011). *Open Data Guidelines - Civic Commons Wiki.* [online] Available at: http://wiki.civiccommons.org/Open_Data_Guidelines/

♦ Wiki.civiccommons.org (n.d). *Open Standards Policy - Civic Commons Wiki.* [online] Available at: http://wiki.civiccommons.org/Open_Standards_Policy/

♦ Tauberer, J. (2014). *U.S. Federal Open Data Policy - Open Government Data: The Book*. [online] Available at: https://opengovdata.io/2014/us-federal-open-data-policy/

♦ Open Data Institute (n.d). Publisher's Guide to Open Data Licensing. [online] Available at: https://theodi.org/guides/publishers-guide-open-data-licensing

♦ Sunlight Foundation (n.d). *Open Data Policies At Work*. [online] Available at: https://sunlightfoundation.com/policy/opendatamap/

Civil Society Organization (CSO) working in Nepal

♦ Kathmandu Living Labs: http://kathmandulivinglabs.org
♦ Open Nepal: http://opennepal.net
♦ Open Knowledge Nepal: http://oknp.org
♦ Freedom Forum: http://freedomforum.org.np
♦ Code for Nepal: http://codefornepal.org
♦ Accountability Lab: http://accountabilitylab.org
♦ Bikas Udhyami: http://bikasudhyami.com

3.1. Open Data Sources of Nepal

Although the idea of open data entered Nepal in early 2013, the government still lacks proper planning and strategy to move the open data movement forward. There has been a significant improvement in a number of governmental bodies publishing their data online. However, the published data is still not available in open format (most of the data are published in PDF format). Similarly, the use of non-open licenses while publishing data by government has been a great hindrance for Nepal.

In Nepal's context, open government data is aligned with the Right to Information (RTI) Act in numerous ways. Both RTI and open data facilitate the availability of government data, it extends the availability of data on request to proactive disclosure of government data publicly, preferably over the Internet. Despite having rights to demand and obtain information through the RTI Act with any public and government organization, the Act does not a have legal provision to pressurize public agencies to open up their data [27].

Recommended Data Sources
- **Some of the notable government data sources are**
  - Official Portal of Government of Nepal: https://www.nepal.gov.np
  - National Planning Commission: http://www.npc.gov.np/en
  - Central Bureau of Statistics: http://cbs.gov.np/home
  - Ministry of Finance: http://mof.gov.np/en/
  - Nepal Rastra Bank: https://www.nrb.org.np
  - Ministry of Home Affairs: http://www.moha.gov.np/home
  - Ministry of Education: http://www.moe.gov.np
  - Ministry of Health: http://mohp.gov.np
  - Election Commission Nepal: http://election.gov.np
  - Office of Company Registrar: http://ocr.gov.np

- **Some of the notable international data sources are**
  - World Bank: https://goo.gl/wFjYgH
  - United Nations: https://goo.gl/UGUoCh
  - UN Digital Repository in Nepal: http://un.info.np
  - UNICEF Nepal: https://goo.gl/QwHhJL
  - World Food Programme: https://goo.gl/2sG2aS

- **Some of the notable CSO data sources are**
  - Open Nepal: http://data.opennepal.net
  - Election Nepal: http://electionnepal.org
  - Nepal in Data: https://nepalindata.com
  - NepalMap: http://nepalmap.org

---

[27] Article 19 (2015). *Country Report: The Right to Information in Nepal.* [online] Available at: https://www.article19.org/resources.php/resource/38205/en/country-report:-the-right-to-information-in-nepal [Accessed 22 Sep. 2017].

3.2. Publishing Data in Nepal

Globally different countries have their official open data portals and journals in which they publish the data in a machine-readable format and encourage citizens to reuse those data to build innovative products, stories, and visualization. Since Nepal don't have any kinds of open data policies and portal, different bodies of Nepal government published and share data through their existing websites. However,t most of the published datasets are in PDF format, which isn't machine readable.

If you are working with data in Nepal and looking to publish the data which you had collected or analyzed, we suggest three different ways of data publishing:

1. **Using own website** - you can create a separate space for Data on your website, where you can share the data.

2. **Independent publishing** - you can use existing tools and cloud spaces like Github, Google Drive, DropBox etc for sharing.

3. **Existing Data portals** - you can use the data portals like the Open Nepal Data Portal and Open Knowledge Nepal Data Hub run by civil society organizations in Nepal for publishing and sharing.

As mentioned before, for your data to be open, make sure to provide an open license and attribution to your datasets, so that other users can reuse and add value to the created data without restrictions.

Recommended Publishing Medium
♦ **Existing Data Portals**
  ♦ Open Nepal Data Portal: http://data.opennepal.net
  ♦ Open Knowledge Nepal DataHub: https://old.datahub.io/organization/nepal

♦ **Independent medium**
  ♦ GitHub: http://github.com
  ♦ Google Drive: http://drive.google.com
  ♦ DropBox: http://dropbox.com

3.3. Open Data Stories from Nepal

Despite the slow progress made by Nepal in the open data field, the success stories of open data initiatives launched and led by different governmental bodies and civil society organizations is also increasing. The government is promoting the use of technology in their daily works to engage citizens in decision-making and policy formation. Citizen engagement projects like Hello Sarkar and Hamro police Mobile App are generating more data which are linked with the daily life of citizens. There are also a couple of government bodies like the Ministry of Finance, the Office of Company Registrar and the Election Commission Nepal, who are sharing the collected data in open formats.

Some open data and citizen engagement projects initiated by the Government of Nepal are:

1. **Aid Management Portal** - Nepal's Aid Management Portal provides public access to information on all donor-funded development projects in Nepal. Users can find out what development partners are working on and what is being achieved, read profiles of donors and sectors, use the interactive map to see exactly where projects and programmes are located, and view in-depth reports on a wide range of aid trends.

2. **OCR Open Data** - Office of Company Registrar Nepal opened the data regarding company registration from 2002 till 2072 BS for download and reuse in CSV and XML format. They embarked the Open Government Data (OGD) initiative to increase access to company data and wanted to create an ecosystem of OGD to increase the usability of data.

3. **Hello Sarkar** - Hello Sarkar is a complaint-hearing government mechanism at the Office of the Prime Minister and Council of Ministers. People may connect with Hello Sarkar anytime either via phone call, message or social media to lodge their complaints; they may inform about corruptions and irregularities, criminalities, reliefs and rescues, disasters and other incidents of public concerns.

Civil society organizations and private sector are also using different types of data like Agriculture [28], Health [29], Election [30] etc., to innovate, use and increase the sharing of data in Nepal.

[28] Open Nepal (2017). *Using agricultural data to guide government's food security efforts in Nepal.* [online] Available at: http://opennepal.net/blog/using-agricultural-data-guide-government's-food-security-efforts-nepal [Accessed 2 Oct. 2017].

[29] Open Nepal (2017). *Using supply-chain data to improve the quality of public health services in Nepal.* [online] Available at: http://opennepal.net/blog/using-supply-chain-data-improve-quality-public-health-services-nepal [Accessed 2 Oct. 2017].

[30] Open Nepal (2017). *Online collaboration by Nepal's data community yields a portal for opening local election data.* [online] Available at: http://opennepal.net/blog/online-collaboration-nepals-data-community-yields-portal-opening-local-election-data [Accessed 2 Oct. 2017].

Some open data projects initiated by CSOs/private sector are:

1. **NepalMap** - NepalMap is Code for Nepal project that aims to make data on Nepal more accessible and understandable by creating user-friendly data visualizations on key demographic issues. NepalMap is a friendly and reliable resource for journalists who want to use data to tell stories or discover stories that are of public interest.

2. **Nepal In Data** - Nepal in Data (NiD) is an open data and statistics portal developed by Bikas Udhyami, to make development data and statistics on Nepal from 1950 to present available and accessible to wide audience. The portal aims to serve as a single point of entry from where users can access, interact with, compare and use data and statistics relating to different aspects of Nepal's development.

3. **QuakeMap** - After a massive 7.8 Mg earthquake on April 25, 2015, Kathmandu Living Labs deployed QuakeMap.org with the aim of bridging the information gap so that rescue and relief efforts can be informed and expedited by reports from the ground. QuakeMap.org used crowdsourcing to collect the information about what helps were needed where, who was able to provide them and what other helps had already reached that place.

Recommended Stories
- Aid Management Portal: http://amis.mof.gov.np/portal/
- OCR Open Data: http://www.ocr.gov.np/index.php/en/data/o-g-d
- Hello Sarkar: http://gunaso.opmcm.gov.np/
- Hamro Police Mobile App: https://play.google.com/store/apps/details?id=app.thirdpoleconnect.crs
- Citizen Helpdesk: http://citizenhelpdesk.org/
- Open Nepal: http://opennepal.net/
- NepalMap: https://www.nepalmap.org/
- Nepal in Data: https://nepalindata.com/
- Election Nepal: http://electionnepal.org
- QuakeMap: http://www.kathmandulivinglabs.org/projects/quakemaporg

---

4. Process of working with Data

4.1. Data Cleaning

To increase the usability and reach, published data needs to be cleaned so that it can easily be understood by everybody. Because of the errors, duplicity, and format/standard inconsistencies, data cleaning is hard and most of the time it depends on the type of data. Plan, Analyze/Clean, Implement Automation, Append Missing Data, and Monitor are five simple steps of data cleaning which are used by many to get started. The basic tools and language used for data cleaning are:

- Spreadsheets
- OpenRefine
- Python

There are five key elements to cleaning data - it needs to be valid, accurate, complete, consistent and uniform. The following sections detail the five key elements of clean data:
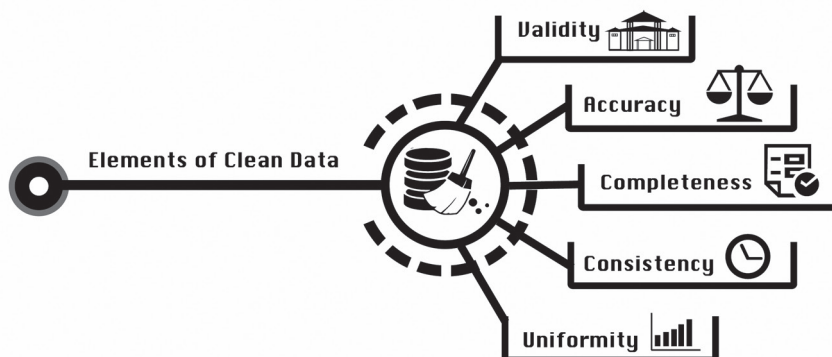
Validity

The precise and exact results acquired from the data collected all depends upon the data validity. Data should be correct and reasonable. Technically, validity can lead to proper and correct conclusions. Errors in data validity are caused by human beings, usually, data entry personnel who enters the data or metadata in different formats.

Accuracy

To be correct and accurate, data must be both the right value and be represented in an unambiguous form, whether data values are correct or not. For example, let's consider a database that contains names, dates of birth, addresses, phone numbers, and email addresses of physicians in Kathmandu. But the value of Date of Birth (DOB) is presented in two different formats, one in USA format (December 17, 2016) and another is European format (12/17/2016). In this case, the conversion from one standard to another causes the accuracy problem and error may occur in the whole

dataset.

Completeness

Completeness of data is mostly two-folded: for all of the records you have do you have all of the data you need? Moreover, do you have all the records that you need? Is all the requisite information available? Are data values missing, or in an unusable state? In some cases, missing data is irrelevant, but when the information that is missing is critical to a specific business process, completeness becomes an issue.

Consistency

Data must be available online at the same URL for the constant amount of time so that others can use the data to show them in different visual and interactive ways without making any changes in the available structure. Fixing inconsistency of data requires a variety of strategies - e.g. deciding which data was recorded more recently, which data source is likely to be most reliable, or simply trying to find the truth by testing both data items.

Uniformity

Uniformity is the degree to which set data measures are specified using the same units of measure in all systems. In a pool of datasets from different locales, all data should follow the same standards. For example, the weight may be recorded either in pounds or kilos, and using different kinds of measures may cause accuracy and uniformity errors.

Recommended Readings

- Salesify (2017). *Keeping it Clean: The Five Step Data Cleansing Process.* [online] Available at: http://www.salesify.com/keeping-it-clean-the-five-step-data-cleansing-process/
- School of Data (2013). *Cleaning up Data Scraped from the Web.* [online] Available at: https://schoolofdata.org/handbook/recipes/cleaning-data-scraped-from-the-web/
- Berkeley Advanced Media Institute (n.d). *Intro to cleaning data.* [online] Available at: https://multimedia.journalism.berkeley.edu/tutorials/cleaning-data/

Guide

- School of Data (2013). *Using Spreadsheets: Spreadsheet Formulae*. [online] Available at: https://schoolofdata.org/handbook/recipes/formulae-with-spreadsheets/
- School of Data (2013). *Using OpenRefine: "Cleaning Data with Refine*. [online] Available at: https://schoolofdata.org/handbook/recipes/cleaning-data-with-refine/
- Stanford University (n.d). *Python: Python tutorials on cleaning and scraping data.* [online] Available at: https://ssds.stanford.edu/python-tutorials-cleaning-and-scraping-data

4.2. Data Extraction

Data extraction is the process of retrieving data out of non-machine-readable or unstructured data sources to further process and analyze the data, or to store it in a raw format. Some sources of unstructured data are: web pages, emails, PDF documents, scanned documents, and so on.

The internet, for example, has a large pool of information. However, the information (or data) on the internet is not structured in the form of datasets and it is difficult to retrieve. Similarly, even with PDFs, the data is available, but you cannot further process and analyze it due to lack of the ability to compare and access the raw data. By extracting such data, you will get access to a whole set of structured data to base your studies and research on, which you can reuse for various purposes.

Extracting data from PDF can be done with:
• Word/Excel converters to extract text from PDF: *https://www.pdftoexcelonline.com*
• Programming, with some libraries existing for Python, Java, and the command line.
• Using Tabula - an offline open-source software specifically designed to get data out of PDF documents.

Hands on with Tabula
Tabula is available for Windows, Mac, and Linux, all you need to do is download and install Tabula along with Java runtime environments. Download Tabula from here: *http://tabula.technology/*. Like every other software Tabula also has some shortcomings and limits. For example, It does not work on merged cells and it cannot detect a scanned PDF document.

Installing Tabula

• Download Tabula and extract the zip file in the desired location.
• Go into the extracted folder to run Tabula program - you will also find Readme in the same folder, which you can use if you are having problem with running.
• It will automatically open in your default web browser.
• If it does not launch on your browser, use this URL - *http://localhost:8080*
• You should now see the user interface of Tabula.

Using Tabula to extract data
• Run Tabula on your browser and click on the Browse button user interface as highlighted on the image and Import the document you want to extract.
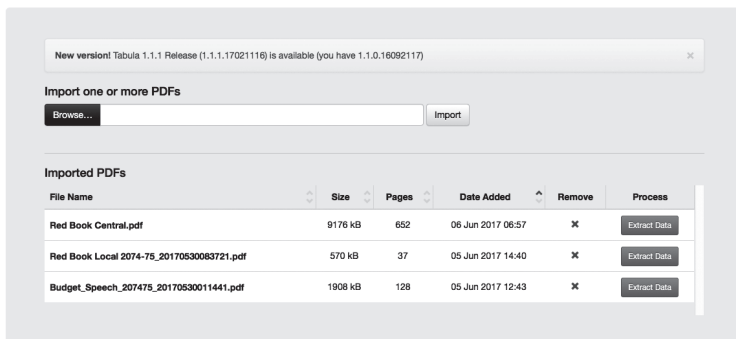
- Once your document is Imported, you need to select the section of the table you want to extract. You can always adjust and repeat your selection for whole documents.
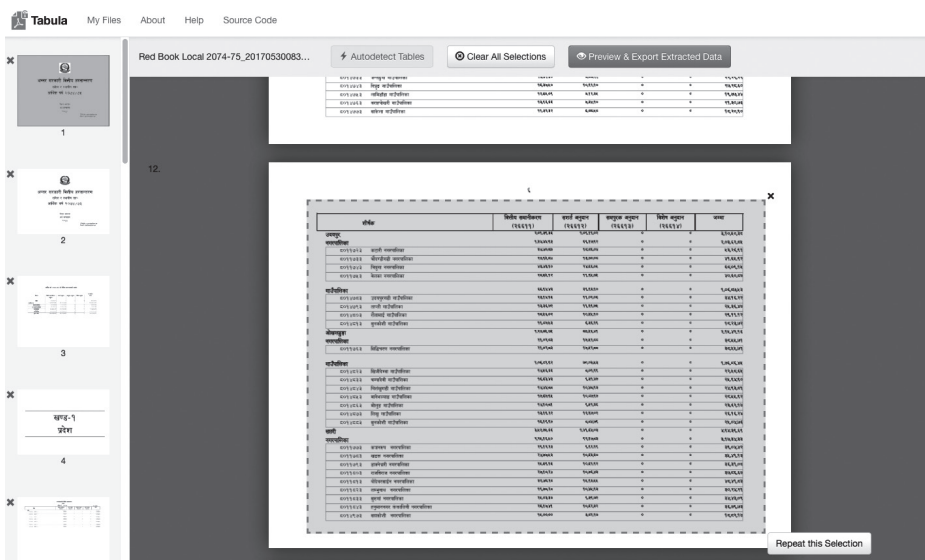
- After making a selection, your data will be processed and extracted by Tabula. Then you will have an option to Export CSV file of the extracted data into your desired location, which can be opened by any application.

Open Data Manual - Compiled by Open Knowledge Nepal

Is the extracted data incorrect?

You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells

Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

← Revise selection(s)

Choose Alternate Extraction Method

The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice** method instead.

▦ Stream   ▦ Lattice

Stream looks for *whitespace* between columns, while Lattice looks for *boundary lines* between columns.

Still look wrong?

Contact the developers and tell us what you tried to do that didn't work.

Red Book Local 2074-75_20170530083...   **Export Format:** CSV   ⊕ Export   Copy to Clipboard

Preview of Extracted Tabular Data

		व	ीय समानीकरण	सशत अनुदान	समपुरक अनुदान	वशष अनुदान	जम्मा
	शीर्षक						
			(26611)	(26612)	(26613)	(26614)	
उदयपुर			2,01,39,36	1,09,11,02	0	0	3,10,50,38
नगरपा लका			1,34,44,93	69,17,92	0	0	2,03,62,85
8011723	कटार नगरपा लका		34,40,87	18,86,04	0	0	53,26,91
8011733	चौदण्डीगढ नगरपा लका		27,95,84	13,70,08	0	0	41,65,92
8011743	यगा नगरपा लका		43,43,10	24,66,05	0	0	68,09,15
8011753	वलका नगरपा लका		28,65,12	11,95,75	0	0	40,60,87
गाउपा लका			66,94,43	39,93,10	0	0	1,06,87,53
8014783	उदयपरगढ गाउपा लका		23,14,16	11,02,06	0	0	34,16,22
8014793	ताप्ल गाउपा लका		13,36,72	11,99,75	0	0	25,36,47
8014803	रौतामाई गाउपा लका		18,56,02	10,55,10	0	0	29,11,12
8014813	सनकोशी गाउपा लका		11,87,53	6,36,19	0	0	18,23,72
ओखलढ ला			1,27,85,75	87,56,41	0	0	2,15,42,16
नगरपा लका			21,02,83	17,52,88	0	0	38,55,71
8011763	स चरण नगरपा लका		21,02,83	17,52,88	0	0	38,55,71
गाउपा लका			1,06,82,92	70,03,53	0	0	1,76,86,45

Recommended Readings

♦ School of Data. (2013). *Scraping websites using the Scraper extension for Chrome.* [online] Available at: https://schoolofdata.org/handbook/recipes/scraper-extension-for-chrome/

♦ School of Data (2013). *Extracting data from PDFs using Tabula.* [online] Available at: https://schoolofdata.org/extracting-data-from-pdfs/

♦ School of Data. (2013). *Making data on the web useful: scraping.* [online] Available at: https://schoolofdata.org/handbook/courses/scraping/

♦ School of Data. (2013). *Scraping websites using the Scraper extension for Chrome.* [online] Available at: https://schoolofdata.org/handbook/recipes/scraper-extension-for-chrome/

♦ Hitchhiker's Guide. (n.d). *HTML Scraping - The Hitchhiker's Guide to Python*. [online] Available at: http://docs.python-guide.org/en/latest/scenarios/scrape/

♦ Classic.scraperwiki.com. (n.d.). *Documentation / Datastore copy & paste guide.* [online] Available at: https://classic.scraperwiki.com/docs/python/python_datastore_guide/

♦ Hirst, T. (2013). *Liberating HTML Data Tables.* [online] School of Data. Available at: https://schoolofdata.org/handbook/recipes/liberating-html-tables/

Tools:

♦ **Basic scraping tools**
  ♦ Tabula: http://tabula.technology/
  ♦ Import.io: https://import.io/

♦ **Extracting data with Python**
  ♦ BeautifulSoup: https://pypi.python.org/pypi/beautifulsoup4
  ♦ Python Mechanize: https://pypi.python.org/pypi/mechanize/
  ♦ Scrapy: http://scrapy.org/

♦ **Web scraping tools**
  ♦ ParseHub: https://www.parsehub.com/
  ♦ ScraperWiki: https://scraperwiki.com/
  ♦ OutWit Hub: https://goo.gl/1Axk88
  ♦ Scraper: https://goo.gl/DZRGZd

4.3. Data Analysis

Data analysis is the process of examining and exploring datasets in order to generate information by cleaning and transforming the data according to your specific needs. Analyzing the data will help you to discover useful information and find the right patterns to represent data into a visual format. There are multiple approaches of doing data analysis according to our data and work types. Segregating types of variables and structure in data is always necessary to make sure the data is accurate and complete. The data structure keeps the data in array, and the file, the record and the table in an organized format. While keeping the data in a certain structure during analysis, we need to care about the types of variables. There are two types of variables, numerical and categorical.

Categorical Variables are made up of a group of categories. For example, Sex (male/female) or Quality (good; bad; average).

Numerical Variables are numbers. They can be counts (e.g. number of participants at a training) or measures (e.g. height of a child).

Name	Sex (Categorical Values)	Age (Numerical Values)
Person 1	Female	23
Person 2	Male	33

Normally two type of analysis are used while playing with data: Qualitative Analysis *generally focuses on descriptive data* and Quantitative Analysis *generally focuses on numerical data.* Both have their own uses and standards. Qualitative Analysis is used when we need to analyze data that is subjective and not numerical. Unlikely, in the quantitative analysis, the data is analyzed through statistical means. There are different offline and online tools, which are used for data analysis. Some of them are:

- Tableau Public
- OpenRefine
- RapidMiner
- Google Fusion Tables
- KNIME
- Import.io

Recommended Readings

- School of Data (2013). *Geocoding Data in a Google Docs Spreadsheet.* [online] Available at: https://schoolofdata.org/handbook/recipes/geocoding/
- Kathryn Lindholm-Leary & Gary Hargett (n.d). [PDF] Available at: http://www.cal.org/twi/EvalToolkit/appendix/toolkit13_sec9.pdf
- Qualitativedataanalysis.net (n.d). *Qualitative and Quantitative Data Analysis.* [online] Available at: http://www.qualitativedataanalysis.net/qualitative-and-quantitative-data-analysis/
- School of Data (2013). *Online geocoding.* [online] Available at: https://schoolofdata.org/handbook/courses/geocoding/

Tools

- Tableau Public: https://public.tableau.com/s/
- OpenRefine: http://openrefine.org
- RapidMiner: https://sourceforge.net/projects/rapidminer/
- Google Fusion Tables: https://goo.gl/XEFUVB
- KNIME: https://www.knime.com
- Import.io: https://www.import.io

4.4. Data Visualization

Data visualization is the presentation of data in a pictorial and graphical format. It helps viewers to see and interpret patterns and trends in the data easily. It. Data visualization is very helpful in capturing a viewer's attention and keeping them engaged throughout the storytelling. Studies show that humans respond to visuals better than any other type of stimuli. The human brain processes visual information 60,000 times faster than text [31].

Data visualization is important when you want to answer business questions, analyze validated data, tell a compelling story, understand information quickly to strategize for the next steps or share your insights with the rest of the team.

Visualization takes care of a complex problem that could be easily ignored and simplifies it using graphs, patterns, and design. It is useful in converting real-time data into rich reports and visualizations with various tools. Today's world contains an unimaginably vast amount of data. Different type of data visualization tools help you make sense all of it. It enables you to look at data differently to discover new answers, make decisions and insights by:

• Telling a Visual Data Story
• Recognize the Signals from the Noise and Colors
• Master a Growing Volume of Data

Hands on with Datawrapper.de

You can use basic vertical bar charts, pie charts, line charts, and scatter plots for easy visualization and storytelling. Please refer to the recommended reading to find out how to visualize data using different kinds of tools. For quick and fast visualization, you can also use an online visualization application such as Datawrapper.de. It can be used to quickly create and publish web-based charts. Datawrapper is helping thousands of journalists on in the world to quickly tell great, data-driven stories.
Steps of using datawrapper.de:
• Open datawrapper.de in your web browser.
• Sign-up for the new account or Login if you already have one.
• On the datawrapper homepage click on create a chart or create a map.

---

[31] Thermopylae Sciences (2014). *Humans Process Visual Data Better.* [online] Available at: http://www.t-sciences.com/news/humans-process-visual-data-better [Accessed 22 Sep. 2017].

- Upload your data from the computer or copy-paste your data.

- Check and describe more about your data.
- Then choose the best way to represent your data.

Give Title, Description and source of your data.

Publish and Embed you visualization.

1 Upload Data ✔ | **2** Check & Describe ✔ | **3** Visualize ✔ | **4** Publish & Embed

What do you want to do with your chart?

The best and easiest way to use a Datawrapper chart is to embed it directly into your website or CMS. You can also download your chart as an image or PDF file to use it in different contexts.

☁
Embed chart on website

< Click here if you want to embed your chart in your **website or CMS**.

or alternatively...

Y
Duplicate this chart

Nepal

Bahun 3	226
Newar 1	887
Tharu 1	737
Magar 1	539

Created with Datawrapper

Recommended Readings

♦ Sas.com (2017). *Data Visualization: What it is and why it matters*. [online] Available at: https://www.sas.com/en_us/insights/big-data/data-visualization.html

♦ Knaflic, C. (2017). *Blog - how would you show this data?* [online] Available at: http://www.storytellingwithdata.com

♦ School of Data (2013). *Creating Line Charts.* [online] Available at: https://schoolofdata.org/handbook/recipes/line-charts-from-spreadsheets/

♦ School of Data (2013). *Walkthrough: Scatterplot.* [online] Available at: https://schoolofdata.org/handbook/recipes/scatterplots-from-spreadsheets/

♦ School of Data (2013). *Creating a Choropleth map.* [online] Available at: https://schoolofdata.org/handbook/recipes/choropleth-maps-from-spreadsheets/

♦ Aisch, G. (2017). *Data Wrapper Tutorial - School of Data Journalism - Perugia.* [online] Available at: https://schoolofdata.org/2013/04/27/data-wrapper-tutorial-gregor-aisch-school-of-data-journalism-perugia/

Tools:
● Non-Developers Visualization Tools
  ♦ Datawrapper: https://datawrapper.de/
  ♦ Infogram: https://infogr.am/en
  ♦ Tableau Public: https://public.tableau.com/
  ♦ Raw: http://raw.densitydesign.org/
  ♦ Plotly: https://plot.ly/
  ♦ Timeline JS: http://timeline.knightlab.com/
  ♦ ChartBlocks: http://www.chartblocks.com/
  ♦ Plotly: https://plot.ly/

● Developers Visualization Tools
  ♦ D3.js: http://d3js.org/
  ♦ FusionCharts: http://www.fusioncharts.com/
  ♦ Chart.js: http://www.chartjs.org/
  ♦ Google Charts: https://developers.google.com/chart/
  ♦ Highcharts: http://www.highcharts.com/

● Map-Based Visualization Tools
  ♦ Leaflet: http://leafletjs.com
  ♦ Mapbox: https://www.mapbox.com
  ♦ CARTO: https://carto.com

4.5. Publishing Data

Data publishing is defined as the process of releasing data in a published form for use and reuse by others. So, this practice consists of the process of preparing datasets and making them available for the public and everyone willing to use it as they wish. Publishing will ensure that data will be considered as a research output that will be available, peer-reviewed, citable, easily discoverable and reusable. These mechanisms will facilitate data transparency and scrutiny and will be used by researchers to increase their academic status, thereby providing an incentive for them to archive and document their data appropriately. However, published data or online data is not necessarily open data. For example, data that is published as an excel table within a PDF document, without an open license, is not open data, because it cannot be easily managed or reused.

To be completely open, the data needs to be available in machine-readable formats. Some of the mostly used open data formats are: JavaScript Object Notation (JSON), Extensible Markup Language (XML), Resource Description Framework (RDF), Spreadsheets, Comma Separated Value (CSV) and Plain Text.
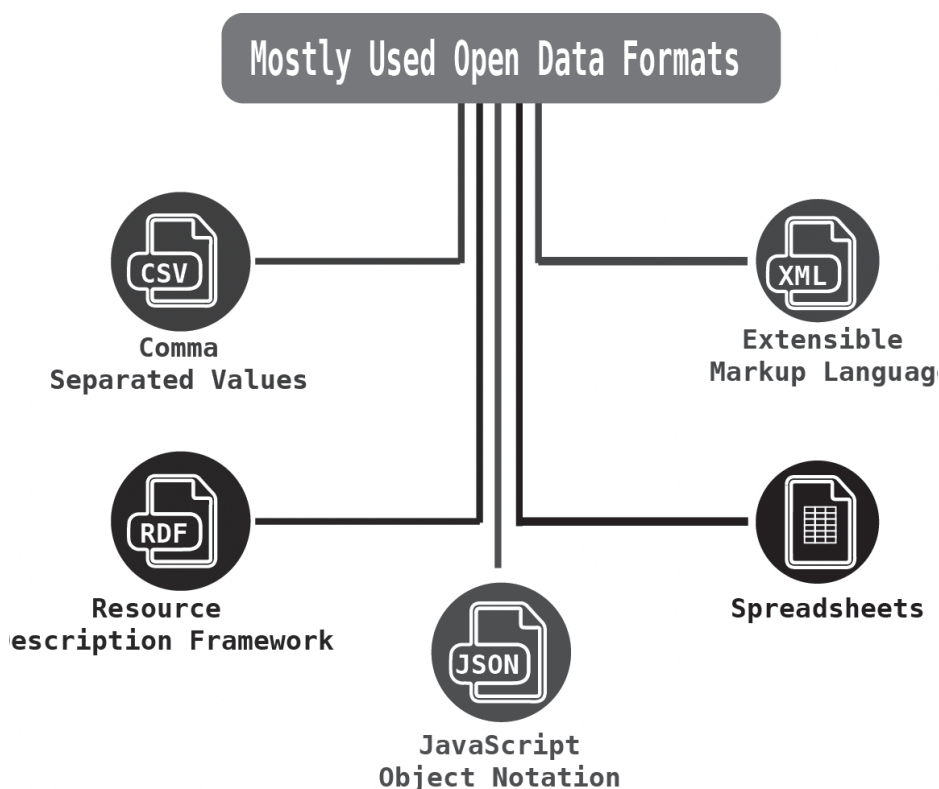
To publish data, data portals and data journals can be used. The most basic definition of an open data portal is, "*a list of datasets with pointers to how those datasets can be accessed*" [32]. Data portals rarely place any restrictions on the type of data that is catalogued or the means by which data is accessed. However, portals offer additional capabilities for both the end user and the publisher. File storage, additional curation tools, integrated data stores are some of the must-have features of data portals. There are a number of open source and commercial data stores, including CKAN, Socrata, and OpenDataSoft which offer a mixture of the outlined features.On the other hand, data journals, unlike traditional journals, focus on the data with a purpose is to expose datasets, rather than producing an extensive analysis and contents. They seek to promote scientific accreditation and reuse, improve transparency of scientific method and results, support good data management practices and provide an accessible, permanent and resolvable route to the dataset.

Open Data Licensing and Formats

There are lots of aspects to openness, but at its most fundamental, the key is how the data is licensed. Data that does not explicitly have an open license is not open data. An open license is one that places very few restrictions on what anyone can do with the content or data that is being licensed. An open license allows others to do things like: republish the content or data on their own website, derive new content or data from yours, make money by selling products that use your content or data, republish the content or data while charging a fee for access.

Creative content, such as text, photographs, slides, and so on, should be licensed using a Creative Commons [33]. Recommended licenses include:

Level of License	Creative Commons License
Public Domain	CC0
Attribution	CC-BY
Attribution & Share-alike	CC-BY-SA

[32] Lost Boy (2015). *What is a data portal?* [online] Available at: https://blog.ldodds.com/2015/10/13/what-is-a-data-portal/ [Accessed 3 Oct. 2017].

[33] Creative Commons (n.d). *Licensing types - Creative Commons*. [online] Available at: https://creativecommons.org/share-your-work/licensing-types-examples/ [Accessed 3 Oct. 2017].

Open Definition have the lists of recommended conformant licenses used by different countries: *http://opendefinition.org/licenses/*

Recommended Readings
- Responsibledata.io (2017). *Responsible Data Handbook | Sharing Data*. [online] Available at: https://responsibledata.io/resources/handbook/chapters/chapter-02c-sharing-data.html/
- Open Knowledge International (n.d). *About - Data Portals.* [online] Available at: http://dataportals.org/about
- Lost Boy (2015). *What is a data portal?* [online] Available at: https://blog.ldodds.com/2015/10/13/what-is-a-data-portal/
- Open Data Institute (n.d). *Publisher's Guide to Open Data Licensing.* [online] Available at: https://theodi.org/guides/publishers-guide-open-data-licensing/
- Opendatahandbook.org (n.d.). *File Formats*. [online] Available at: http://opendatahandbook.org/guide/en/appendices/file-formats/
- greatlearning.in (2016). *Data Storytelling: What It Is, Why It Matters - Great Learning.* [online] Available at: https://www.greatlearning.in/blog/data-storytelling-what-it-is-why-it-matters/

5. Glossary

Accountability

The obligation of an individual or organization to account for its activities, accept responsibility for them, and to disclose the results in a transparent manner. It also includes the responsibility for money or other entrusted property.

Application Programming Interface

A way computer programs talk to one another. Can be understood in terms of how a programmer sends instructions between programs.

Attribution

A license that requires attributing the original source of the licensed material.

Big Data

A collection of large data which cannot be stored, transmitted or processed in traditional ways. The increasing availability and need to process such datasets (for example, huge collections of weather or other scientific data) has led to the development of specialized computer technologies, architectures and programming languages.

Categorical Data

Data that helps put things into categories. E.g.: Country names, Groups, Conditions, Tags.

Citizen Engagement

Actively involving the public in policy and decision-making. Citizen engagement is a central aim of open government, with the aims of improving decision-making and gaining or retaining citizens' consent and support.

Creative Commons

A non-profit organization founded in 2001 that promotes reusable content by publishing a number of standard licenses, some of them open, that can be used to release content for reuse, together with clear explanations of their meaning.

Crowdsourcing

Dividing the work of collecting a substantial amount of data into small tasks that can be undertaken by volunteers.

CSV

Comma Separated Values (CSV) is a standard format for spreadsheet data. Data

is represented in a plain text file, with each data row on a new line and commas separating the values on each row. As a very simple open format, it is easy to consume and is widely used for publishing open data.

Data collection

Datasets are created by collecting data in different ways: from manual or automatic measurements (e.g. weather data), surveys (census data), records of decisions (budget data) or ongoing transactions (spending data), aggregation of many records (crime data), mathematical modeling (population projections), etc.

Data management

The policies, procedures and technical choices used to handle data through its entire lifecycle from data collection to storage, preservation, and use. A data management policy should take account of the needs of data quality, availability, data protection, data preservation, etc.

Data wrangler

A person converting data into usable form so that they can be easily used with automated or semi-automated tools. Data wrangling may include further data cleaning.

Google Chrome

Google Chrome is a freeware web browser developed by Google. It was first released in September 2008, for Microsoft Windows, and was later ported to Linux, macOS, iOS, and Android.

JSON

JavaScript Object Notation (JSON) is a simple but powerful format for data. It can describe complex data structures, is highly machine-readable as well as reasonably human-readable, and is independent of platform and programming language, and is therefore a popular format for data interchange between programs and systems.

Metadata

Information about a dataset such as its title and description, a method of collection, author or publisher, area and time period covered, license, date, and frequency of release, etc.

National Information Commission

National Information Commission (NIC) is an independent body for the implementation of Right to Information as per the provisions of the RTI Act in Nepal. It was established in 2008 and responsible for the protection, promotion, and practice of RTI in

Nepal.

Open Data Barometer
The Open Data Barometer (ODB) aims to uncover the true prevalence and impact of open data initiatives around the world. It analyzes global trends and provides comparative data on countries and regions using an in-depth methodology that combines contextual data, technical assessments, and secondary indicators.

Global Open Data Index
The Global Open Data Index (GODI) is the annual crowdsourced survey that measures the openness of open government data, run by Open Knowledge International.

Optical Character Recognition
Optical Character Recognition (OCR), is a technology that enables you to convert different types of documents, such as scanned paper documents, PDF files or images captured by a digital camera into editable and searchable data.

Qualitative Data
Qualitative data is data telling you something about qualities: e.g. description, colors etc. Interviews count as qualitative data.

Quantitative Data
Quantitative data tells you something about a measure or quantification. Such as quantity of things you have, the size (if measured) etc.

Scraping
Extracting data from a non-machine-readable source, such as a website or a PDF document, and creating structured data from the result. Screen-scraping a dataset requires dedicated programming and is expensive in programmer time, so is generally done only after all other attempts to get the data in structured form have failed.

Share-alike
A license that requires users of a work to provide the content under the same or similar conditions as the original.

Socio-economic data
Socio-economic data is data about human's, activities, and space and/or structures used to conduct human activities. The specific classes of socio-economic data include demographics, housing, migration, transportation, economics, retailing etc.

Tabula

If you have ever tried to do anything with data provided to you in PDFs, you know how painful it is - there is no easy way to copy-and-paste rows of data out of PDF files. Tabula allows you to extract that data into a CSV or Microsoft Excel spreadsheet using a simple, easy-to-use interface. Tabula works on Mac, Windows, and Linux.

Taxonomy

Taxonomy refers to the hierarchical classification of things. One of the best known is the Linnaean classification of species - still used today to classify all living beings.

Version-tracking

Developers may wish to compare today's version of some software with yesterday's or older one. Since version control systems keep track of every version of the software, this becomes a straightforward task. Knowing the what, who, and when of changes will help with comparing the performance of particular versions, working out when bugs were introduced, and so on.

About The Manual

The Open Data Manual is especially focused on Nepal's young generation of digital natives, who want to know more about the background, history, sources, stories and process to extract, analyze, clean and visualize the available data in Nepal. The manual will make the Nepali youth more aware of open data's benefits, will fill in the gap of data education, and will better prepare young people for a rapidly changing data scenario. The Open Data Manual is compiled by Open Knowledge Nepal from a series of online resources covering key data topics to explain open data, its benefits, and technicalities as a reference and recommended guide for university students, the private sector, and civil society.

About Open Knowledge Nepal

Open Knowledge Nepal is an open network of open knowledge enthusiasts. We are a non-profit organization comprised of openness aficionados - mainly self-motivated youths who believe that openness of data is powerful in order to have a participatory government with civil society, eventually leading to sustainable development. We have been a working group of Open Knowledge International since February 2013.

OPEN KNOWLEDGE NEPAL

✉ openknowledgenepal@gmail.com

🌐 http://oknp.org