



Privacy and the limits of Open Data in South Africa

Nobubele Angel Shozi and Anwar Vahed

Council for Scientific and Industrial Research (CSIR), Pretoria,
South Africa.

31 May 2018

IASSIST and Carto 2018 Conference, Montreal.

Presentation Layout



- Once upon a time.....
- What is the purpose of this presentation?
- Open Data defined
- Survey of Current Open Data Repositories in South Africa
- Privacy criteria + Privacy issues in Open Data Repositories of South Africa
- Protection of Personal Information (POPI) Act of South Africa
- POPI Act assessment
- Privacy recommendations
- Conclusion

Once upon a time.....



The Netflix logo is displayed on a solid red rectangular background. The word "NETFLIX" is written in a white, bold, sans-serif font. Each letter has a subtle 3D effect with a dark shadow cast to the right.

Image Sourced from: https://thenextweb.com/apps/2016/06/20/netflix-just-changed-logo/#.tnw_6ny0a083

The AOL logo is shown in a large, bold, black sans-serif font on a white background. The letters "A", "O", and "I" are significantly larger than the "L", which is followed by a period.

Image Sourced from: https://commons.wikimedia.org/wiki/File:AOL_logo.svg



Image Sourced from:
https://commons.wikimedia.org/wiki/File:Target_logo.svg

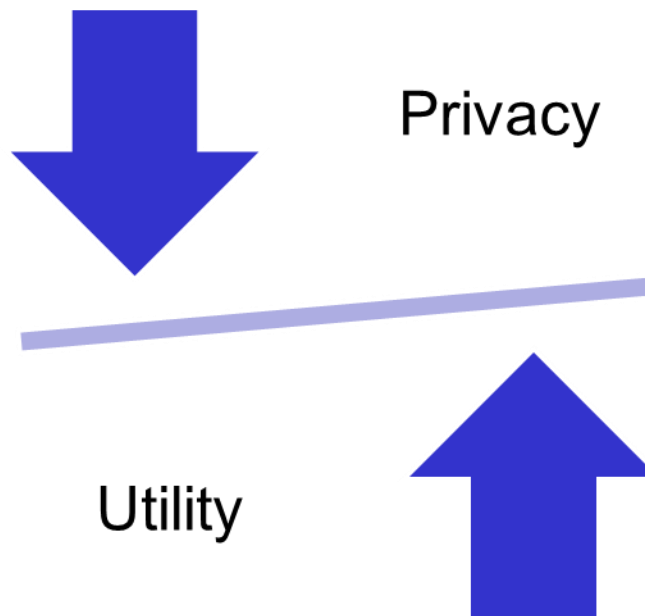
What is the purpose of this presentation?



- Analyse Open Data Repositories in South Africa
- Identify privacy related issues
- Assess repositories privacy policies based on Privacy Law in South Africa
- Provide privacy recommendations

Open Data Defined

- Open data can be defined as data that is “freely used, re-used and redistributed” [1].
- Characteristics: Public domain, Whole set, Readily process able by a computer, Open format, Open licence that allows it to be reuse, redistributed, and Allows for modification.
- Open Data has benefits however there are privacy risks.



Survey of Current Open Data Repositories in South Africa



- Five (5) South African Open Data Repositories were surveyed:
 - South Africa Data Portal
 - Code for South Africa Portal
 - OpenAfrica by Code for Africa
 - South African Cities Open Data Almanac (SCODA Beta)
 - DataFirst



SOUTH AFRICA
DATA PORTAL



[Sign Up](#) [Login](#)

[HOMEPAGE](#)

[DATA CATALOG](#)

[GALLERY](#)

[SDG](#)

Dataset Browser

Can't find what you are looking for? [Suggest a data source.](#)

Filter: [All](#)

[Topics](#) [Sources](#) [Regions](#) [Search](#)

Name: [C](#) [D](#) [H](#) [I](#) [L](#) [M](#) [P](#) [S](#)

- C** CPI (2)
- D** Demographics (1)
- H** Household Expenditures (1)
- I** Index (2)
 - Inflation (2)
- L** Labor (1)
 - Living conditions (1)
- M** Malaria (1)
 - Manufacturing (1)
 - Mining (1)
- P** PPI (1)
 - Prices (1)
 - Production (2)
- S** Sales (2)

South Africa Data Portal



★ Population Statistics of South Africa, 2011

Demographics

✓ Verified

👁 379

★ 4

📄 1

[Embed](#)

▼ Location

Type to filter

- ☐ South Africa
 - ☐ Eastern Cape
 - ☐ Free State
 - ☐ Gauteng
 - ☐ KwaZulu-Natal
 - ☐ Limpopo
 - ☐ Mpumalanga
 - ☐ Northern Cape
 - ☐ North West
 - ☐ Western Cape

► Variable

► Sex

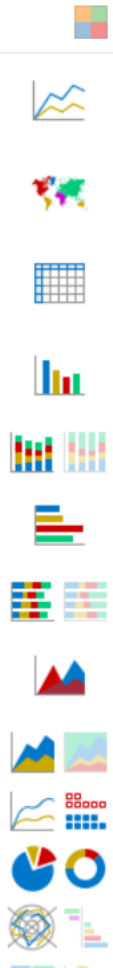
Use dimension filter to select data to visualize. Add/remove elements by selecting/deselecting checkboxes, switch between categories by clicking on their names.

Switch between different visualizations on the fly. Based on your selection recommended gadgets are shown on top with large tiles.

You don't have to start from scratch. Load [random selection](#) or [browse dashboards](#) to see what other people did

Please input your query here

Search



Code for South Africa Portal (OpenUp)



Categories

A-Z

Business
Education
Fun
Government
Health
Personal
Transport

Close

Categories

Business

Education

Fun

Government

Health

Show All...

View Types

Calendars

Charts

Data Lens pages

Datasets

Tags npo, dsd

API Docs

Updated
July 18, 2015
Views
8,187

Police Statistics | 2005 - 2017

Government

Dataset

Police statistics per police station for 29 different crime categories. Includes data from 2005 - 2017. This dataset is based on the data released by SAPS but formatted in a way that allows for proper...

[More](#)

Updated
May 9, 2018
Views
7,303

Tags police, crime

API Docs

Roll-up NGOs

Business

Filtered View

Officers of NPOs in South Africa

Tags No tags assigned

Updated
January 5, 2018
Views
6,756

Code for South Africa Portal (OpenUp)



Take me to the data Help



Sign In

Non-profit database

Scraped from <http://www.dsd.gov.za/npo/> in 2014



Find in this Dataset

More Views

Filter

Visualize

Export

Discuss

Embed

About

id	source_url	status	category	fax
8	http://www.npo.gov.za/PublicNpo/...	Registered	Social Services > Self-help > Sel...	011 485 38
10	http://www.npo.gov.za/PublicNpo/...	Registered	Social Services > Services to Ch...	(011) 4729
13	http://www.npo.gov.za/PublicNpo/...	Registered	Social Services > Services to Ch...	6869320
147,572	http://www.npo.gov.za/PublicNpo/...	Registered	Social Services > Services to Ch...	
17	http://www.npo.gov.za/PublicNpo/...	De-registered	Health > Other Health Services ...	(0325) 816
16	http://www.npo.gov.za/PublicNpo/...	Registered	Law, Advocacy, and Politics > Ci...	012 344 65
14	http://www.npo.gov.za/PublicNpo/...	Registered	Religion > Religious Congregati...	012 37919
147,587	http://www.npo.gov.za/PublicNpo/...	Registered	Social Services > Services to Ch...	
3	http://www.npo.gov.za/PublicNpo/...	Registered	Health > Other Health Services ...	021-886-51
2	http://www.npo.gov.za/PublicNpo/...	Registered	Health > Other Health Services ...	
7	http://www.npo.gov.za/PublicNpo/...	Registered	Education and Research > Rese...	011 489 92
9	http://www.npo.gov.za/PublicNpo/...	Cancelled	Education and Research > Prim...	021 386 29
12	http://www.npo.gov.za/PublicNpo/...	Registered	Education and Research > Oth...	012-328 77
4	http://www.npo.gov.za/PublicNpo/...	Registered	Health > Other Health Services ...	028-72219

More Views

Dataset Snapshots

Views

Sort by

Show only

Search views



ECD Centres

Updated July 18, 2015

Scraped from <http://www.dsd.gov.za/npo/> in 2014



ECD Centres

Updated July 19, 2015

Scraped from <http://www.dsd.gov.za/npo/> in 2014

< Previous Next >

Showing rows 1-50 out of 152,443

OpenAfrica by Code for Africa



[Datasets](#) [Organisations](#) [Groups](#) [About](#) [Data Requests](#)

[Log in](#) [Register](#)

Africa's Largest Volunteer Driven **Open Data** Platform

openAFRICA aims to be largest independent repository of open data on the African continent



57 organisations
curating

Assessment of Varieties of Cassava for high yield, disease resistance in Adva...

Updated May 26, 2018 | Created May 9, 2018

Assessment of Varieties of Cassava for high yield, disease resistance in Advanced Yield Trial (13 clones) in Ibadan 2000/2001 Breeding Season

url

Tanzania Non-Government Organisation

Updated May 22, 2018 | Created May 22, 2018

Code for Tanzania

This dataset has no description

XLSX

AFRICA 2018 POPULATION AND INTERNET USERS STATISTICS

Updated May 18, 2018 | Created March 16, 2018

Code for Tanzania

This dataset has no description

XLSX

Election & Voter Data

Updated May 18, 2018 | Created May 6, 2016

Code for Uganda

Data extracted from Uganda Electoral Commission

CSV

PDF

Unnamed resource

[Go to resource](#)[Data API](#)

URL: <https://africaopendata.org/dataset/bebd75a2-49ff-40bd-b293-0e3f61d873de/resource/5c2df644-9f79-4fd6-a755-644e6151d3ad/download/tanzania-non-government-org...>

[Data Explorer](#)[Embed](#)

Grid Graph Map 1586 records « 1 – 100 »

[Go »](#)[Filters](#)

_id	Name of...	Category	Region	District
1	Arusha ...	Women	Arusha	Arusha
2	SEEWA...	ORPHA...	Arusha	Arusha
3	ACTION ...	VURNE...	Arusha	Arusha
4	KAMAM...	EDUCAT...	Arusha	Arusha
5	FOOD\$...	ENVIRO...	Arusha	
6	ARUME...		Arusha	Arumeru
7	KITUO C...	POWER ...	ARUSHA	

South African Cities Open Data Almanac (SCODA Beta)



SCODA ^{BETA}

HOW-TO

INDICATORS

MAP

ABOUT

EXPLORER

(currently).

SCODA offers an interactive user platform comprising:

The Explorer

The Explorer allows users to download full datasets for the SoCR Almanac as well as other key data sets. Data can be downloaded as a csv file or visualized and downloaded as a PNG graphic file. Over time, all data sets generated by the SACN will be made available to download here. The demographic modelling data can also be downloaded in CSV format.

Explorer

Data visualisations (infographics)

The Indicators button allows users to explore 12 key city indicators as interactive infographics. Each infographic contains an explanation of the indicator as well as reference to where the data was sourced. All of the data for these Indicators is contained within the Almanac

Indicators

Map

The Map highlights the Cities whose data is contained within SCODA as well as key data. The SACN, through a partnership with the CSIR, has also supported the development of more detailed spatial analyses and visualisations of city development which are available at: http://stepsa.org/sacn_socr.html.

Map

[Home](#) > [Open Data Portal](#)

Search by Keyword



Open Data Portal

in study description

in variable description

[Search](#) [Reset](#)





Filter by Year

Show studies conc

1948 and 20

Filter by Data A

☒ All

- ☐  Public use c
- ☐  Licensed da
- ☐  Data enclav
- ☐  Data availab

Filter by Count

☒ All

[View / Select More](#)

Filter by Topic

☒ All

[View / Select More](#)

Select Topics

0 selected | [Clear](#)



☐ DEMOGRAPHY AND POPULATION

- ☐ censuses (7)
- ☐ fertility (17)
- ☐ migration (22)
- ☐ morbidity and mortality (27)

☐ ECONOMICS

- ☐ consumption/consumer behaviour (17)
- ☐ economic conditions and indicators (18)
- ☐ economic policy (1)
- ☐ economic systems and development (1)
- ☐ income, property and investment/saving (26)
- ☐ rural economics (4)

☐ EDUCATION

- ☐ basic skills education (13)
- ☐ compulsory and pre-school education (10)
- ☐ educational policy (6)
- ☐ life-long/continuing education (1)
- ☐ post-compulsory education (6)

[Cancel](#)

[Apply filter](#)

by: Ghana Centre for Democratic Development (GCD), Institute for Justice and Reconciliation in South Africa (IJR, South Africa), Institute for Empirical Research in Political Economy (IREEP,

Created on: Feb 15, 2018 Last modified: Feb 15, 2018 Views: 1889 Downloads: 58



RIA African Mobile Pricing 2010-2015
Africa, 2010-2015

DataFirst

[DataFirst Home](#)
[Open Data Portal](#)
[Open African Data sites](#)
[Citations](#)
[Contact us](#)

[+ SHARE](#)

[Home](#) › [Open Data Portal](#) › [DATA](#) › [KEN-ETCP-TDK-1990-2016-V1](#)

Kenya - Tobacco Data for Kenya 1990-2016



Reference ID ken-etcp-tdk-1990-2016-v1

Year 1990 - 2016

Country Kenya

Producer(s) Data on Alcohol and Tobacco in Afrca Project - Economics of Tobacco Control Project, Southern Africa Labour and Development Research Unit, University of Cape Town

Sponsor(s) International Development Research Centre - IDRC - Funder

Collection(s) [Data on Alcohol and Tobacco in Africa Project](#)

Created on Jun 08, 2017

Last modified Jun 08, 2017

Page views 3006

Downloads 1887

[Documentation](#)
[Study Description](#)

[Get Microdata](#)

Documentation

Download the questionnaires, technical documents and reports that describe the survey process and the key results for this study.

Reports

Economics of tobacco taxation in Kenya



7.26 MB

Tobacco industry interference in Kenya



1.09 MB

Privacy issues in Open Data Repositories of South Africa



- **Access:**
 - How a user gains access to the repository.
 - Create account
 - Verification of account
 - Security of site
- **Dataset:**
 - View/Download rights
 - Upload rights
 - Confidential/Sensitive data
 - Metadata
- **Policy:**
 - What policies are in place?

Privacy issues in Open Data Repositories of South Africa



		South Africa Data Portal	Code for South Africa (OpenUp)	Open Africa by Code for Africa	SCODA	DataFirst
Access	Any user can create account	X	X	X	-	X
	Approval process for new account	-	-	-	-	-
	Https URL	X	X	X	-	X
	Requires strong password for user account	X	X	-		X
Dataset	Anyone allowed to view/download	X	X	X	-	-
	Only registered users allowed to view/download	-	-	-	-	X
	Request required to download	-	-	-	-	X (Sensitive)
	Anyone allowed to upload	-	-	-	-	-
	Only registered users allowed to upload	X	X	-	-	-
	PII found	-	X	X	-	X
Policy	Policies in place	-	X (Privacy, TC's)	X (Privacy, TC's, other)	-	-

Privacy laws – A view

- Privacy has become vital

Privacy has become an urgent issue in data use: In February 2016, President Obama recognized the need for clear guidelines by establishing the Federal Privacy Council.²

Simmons & Simmons

Data Protection landscape

□ African organizations

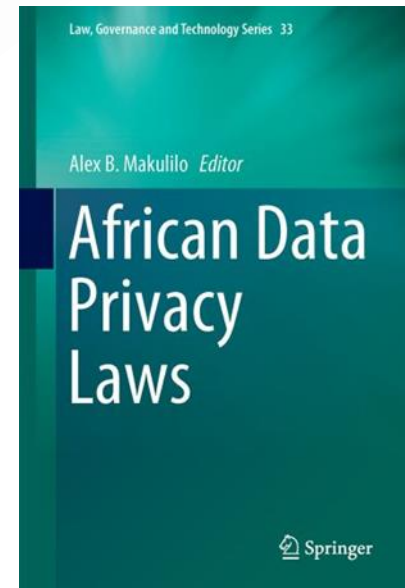
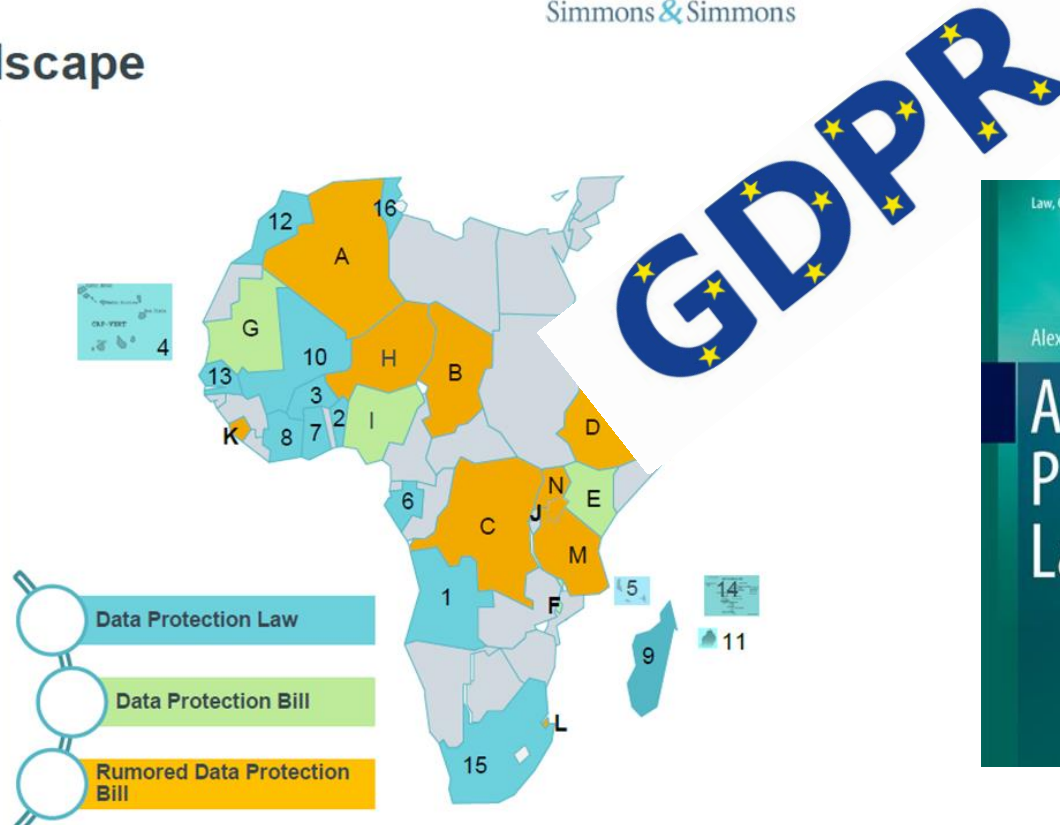
- African Union
- ECOWAS

□ African countries having adopted a DP regulation

- | | |
|-----------------|------------------|
| 1. Angola | 10. Mali |
| 2. Benin | 11. Mauritius |
| 3. Burkina Faso | 12. Morocco |
| 4. Cape Verde | 13. Senegal |
| 5. Comoros | 14. Seychelles |
| 6. Gabon | (Not in force) |
| 7. Ghana | 15. South Africa |
| 8. Ivory Coast | 16. Tunisia |
| 9. Madagascar | |

□ DP Bills (and rumoured bills)

- | | |
|---------------|-----------------|
| A. Algeria | I. Nigeria |
| B. Chad | J. Rwanda |
| C. DRC | K. Sierra Leone |
| D. Ethiopia | L. Swaziland |
| E. Kenya | M. Tanzania |
| F. Malawi | N. Uganda |
| G. Mauritania | |
| H. Niger | |



E.U. Fines Facebook \$122 Million Over Disclosures in WhatsApp Deal

POPI Act of South Africa



- The Protection of Personal Information (POPI) Act of South Africa is a law that has been enacted in South Africa to protect the personal information of South African citizens. The POPI Act has eight (8) conditions that need to be complied with:
 - **Accountability.**
 - **Processing Limitation** - Personal information should only be processed in a manner that ensures that the data subject is protected and only the necessary information is collected for processing.
 - **Purpose Specification** - Information must only be processed for the intended purpose.
 - **Further Processing Limitation** - Any further processing that needs to be done should comply with the initial purpose for the collection of the data.
 - **Information Quality** - Information must always be complete and accurate.
 - **Openness** - The data collector should always be transparent and open to the data subject about data that is collected.
 - **Security Safeguards** - Adequate safeguards to ensure privacy and confidentiality are the responsibility of the data collector/holder.
 - **Data subject participation** - The data subject can be able to ask the data holder to see the data that is held about them.

POPI Act assessment

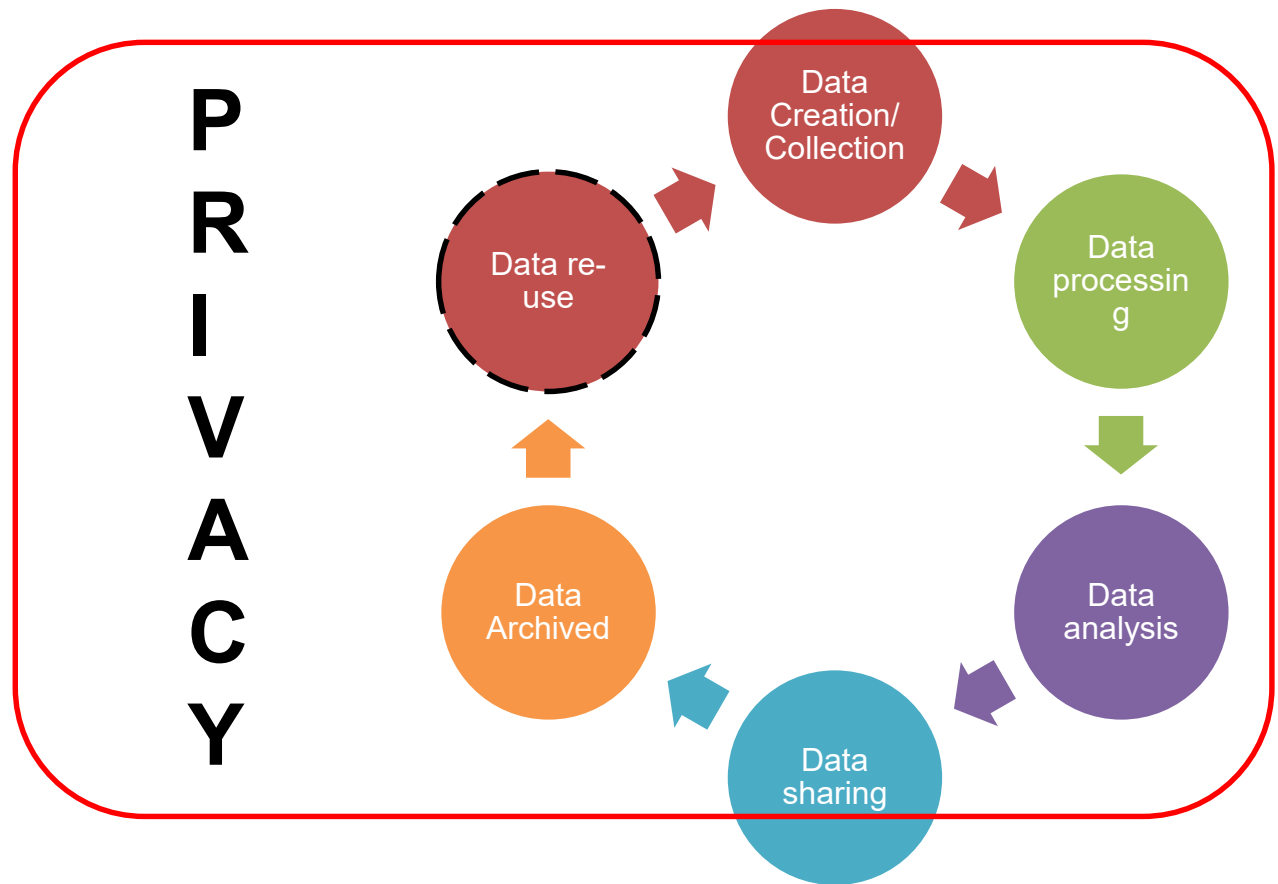


- Only two of the data repositories have privacy policies. Based on the privacy policies of the data repositories, we analyse to see if the POPI act conditions are met:

POPI Criteria	South Africa Data Portal	Code for South Africa (OpenUp)	Open Africa by Code for Africa	SCODA	DataFirst
Accountability	-	X	-	-	-
Processing Limitation	-	X	X	-	-
Purpose Specification	-	X	X	-	-
Further processing limitation	-	X	X	-	-
Information Quality	-	X	X	-	-
Openness	-	X	X	-	-
Security Safeguards	-	X	-	-	-
Data subject participation	-	X	X	-	-

Privacy Recommendations

- Privacy needs to be considered in each step of the data lifecycle (before the data is put in an open data repository).



DISCLAIMER: The recommendations provided are specifically for the repositories

Privacy Recommendations



		Privacy Recommendations
Access	Any user can create account	It is acceptable for any user to create account.
	Approval process for new account	Recommended.
	Https URL	Recommended
	Requires strong password for user account	Recommended
Dataset	Anyone allowed to view/download	Recommended if data is not sensitive.
	Only registered users allowed to view/download	Recommended if data is sensitive.
	Request required to download	
	Anyone allowed to upload	Not recommended.
	Only registered users allowed to upload	Recommended however the data would still need to be verified.
	PII found or requested	PII should be removed from datasets as it infringes on privacy
Policy	Policies in place	Privacy policies that comply with privacy laws of the country are highly recommended.

PII and Privacy?

Privacy Recommendations



- Personally identifiable information such as emails, names, addresses were found in some datasets. The following privacy mitigation techniques are recommended:
 - Removal of fields that contain sensitive data. Removing the fields has a positive impact on privacy but has a negative impact on utility as these records could become useless without certain fields.
 - Aggregating data by providing summarised versions of data. This has a positive impact on privacy. The data however is not very useful as there is no raw data but aggregated data.
 - Generalizing data. The more the generalization, the harder it is to re-identify individuals and the less useful the data is.
 - K-anonymity. Techniques such as K-anonymity can be used to ensure that at least k number of people can be linked to a record.
 - Adding noise to the data. Adding noise ensures more privacy but less utility.
 - Differential privacy is able to strike a more balanced contrast between privacy and utility.

Conclusion



- Privacy and Utility are two issues that seem to be difficult to balance when trying to preserve the privacy of individuals.
- Anonymization is not a 100% safe solution BUT it does provide a certain amount of protection.
- This paper has investigated a number of South Africa data repositories to identify any privacy limits that these open data repositories have.
- A number of privacy limitations were identified and adequate privacy mitigation techniques recommended.
- Future work: This study can be further extended to other African countries and provide an African perspective on open data repositories and the effects of privacy.

Thank you

- Email: nshozi@csir.co.za
- Twitter: @angelshozi

