



# Data Management & Publication with CKAN

Abdul-Hakeem Shaibu  
*@hkmshb*

# Topics Covered

- Introduction
  - ◆ What is CKAN
  - ◆ Who uses CKAN?
- CKAN Features
- Publishing Data on CKAN
- CKAN Architecture
- Experience with CKAN

# Introduction



## What is CKAN?

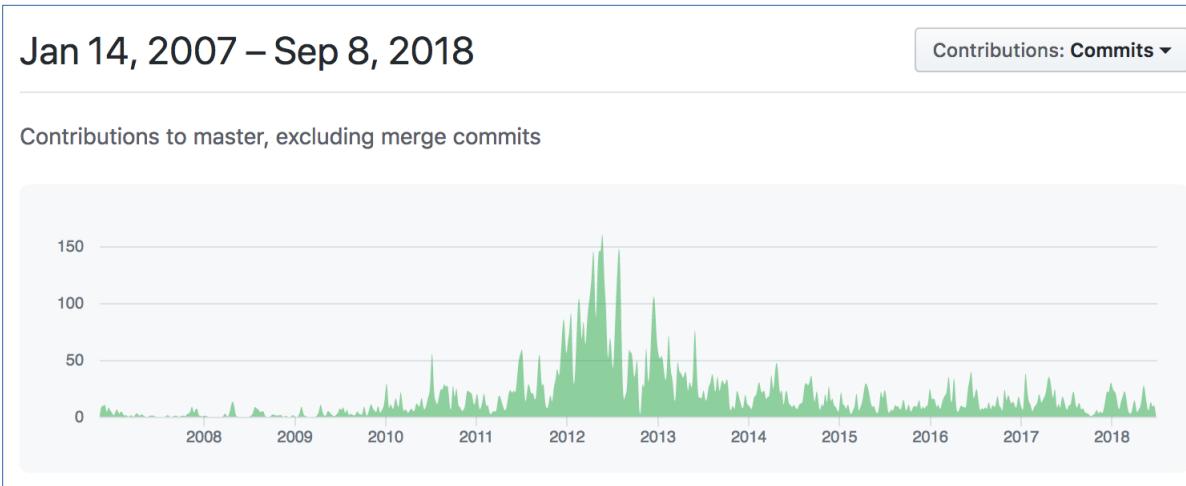
*A data management platform used to manage a curated list of datasets and provides pointers on how those datasets can be accessed.*

**C**omprehensive  
**K**nowledge  
**A**rchive  
**N**etwork

- **open source** software for data catalogs
- written in Python
- GNU Affero General Public License (AGPL) v 3.0  
<https://choosealicense.com/licenses/agpl-3.0/>
- GitHub: <https://github.com/ckan/ckan>

# CKAN Overview (cont.)

Started 2007... Over 10 years in development



Maintained by community of developers

## GitHub Stats.

- 18,851 commits
- 80 releases
- 193 contributors
- 1,000+ forks

# CKAN Overview (cont.)

- Governance: CKAN Association
  - <https://ckan.org/about/association/>
- Association Members



# Who uses CKAN?

CKAN is being used around the world by numerous:

- Governments
- Organizations
- Civil Societies
- Institutions

find more at:

<https://ckan.org/about/instances>



**data.go.jp**  
(third-party submission) This site was set up as a trial version of the Data Catalog Site. Its purposes are to provide a sphere for the use of data owned by different governmental ministries and agencies as open data and to present the image of open data to both data providers and data users.  
Japan



**Data.gov**  
The US government announced in January 2013 that they were moving to CKAN for their open data catalogue, combining datasets previously published in different places. The relaunch went live in May 2013.  
USA



**data.gov.ie**  
Ireland's national open data portal was launched in July 2014. It is operated by the Department of Public Expenditure and Reform.



**data.gov.au**  
Data.gov.au provides an easy way to find, access and reuse public datasets from the Australian Government.  
Australia

# Who uses CKAN?

- Known instances of CKAN used in Africa
- Open data portals available to the general public

The Polio Emergency Operations Center website features a prominent timer at the top right: "Count up from last polio report 749 : 11 : 12 : 24". Below the timer, there's a quote from Dr. Faisal Shuaib: "At this last phase of polio eradication in Nigeria, only a quality, timely and complete data generated from field, analyzed and transformed into meaningful report/result can guarantee us better understanding of actual and real situation for informed decision making." The quote is attributed to Dr. Faisal Shuaib | ED, NHPGDA.

Below the quote, there are sections for Immunization, Surveillance, and GIS. The Immunization section shows coverage data for LGA, State, and National levels. The Surveillance section discusses monitoring and assessing impact. The GIS section provides public access to published GIS data. At the bottom, there are links to datasets, organizations, groups, and resources, along with a "Most recent datasets" section and a "Partners" section featuring logos for WHO, UNICEF, CDC, Rotary, and others.

The GRID3 Nigeria website has a large banner for the "GEO-REFERENCED INFRASTRUCTURE AND DEMOGRAPHIC DATA FOR DEVELOPMENT (GRID3) PROJECT". Below the banner is a map of Nigeria with state boundaries. To the right, there are nine categories of datasets: AGRICULTURE, BUSINESS AND ECONOMY, COMMUNICATION, CRIME AND JUSTICE, EDUCATION, EMERGENCY SERVICES, PARTNERS, RELATED LINKS, and PARTNERS. The RELATED LINKS section includes links to About Portal, Terms of use, Privacy Policy, Newsletter, FAQs, and Contact Us. The PARTNERS section lists the Federal Republic of Nigeria, BILL & MELINDA GATES foundation, and eHealth Africa.

The DATA.Ug website is described as "The home of Uganda's Open Data". It includes a search bar for "Agriculture" and a news item titled "PLE performance trend, 2012, 2013". The news item notes that prior to 2011, the best results were from PLE, but since 2011, the published results from the local news paper have been better. It also mentions the "Data for Development" initiative.

**Open Data in Uganda**

The Africa Open Data website features a map of Africa and a search bar for "Search Your Data". It displays datasets from various countries, including Kenya, Ghana, and Nigeria. The Kenya section shows "Kenya Demographic Health Survey - May 27 to 26, 2013" and "Kenya Downloads". The Ghana section shows "2004 Ghana Budget" and "Ghana Data Site For the Republic of Ghana".

**Africa Open Data**

The Lake Chad Basin Data Portal website features a map of the Lake Chad Basin area. It includes a sidebar with links to Agriculture, Environment, Education, and Management. The main content area shows a large image of people in a boat and a map of the basin.

**LakeChad Basin**

**LAKE CHAD BASIN DATA PORTAL**

# CKAN Features



# CKAN Features

## Publish & Manage Data

- Publish data through
  - ◆ web interface
  - ◆ API, or
  - ◆ harvesting tools
- Similar or related resources are collected together in the same dataset
- Role based User management
- Admin | Editor | Member

Dataset Groups Activity Stream Manage PRIVATE

## Acute Flaccid Paralysis (AFP) Case-Based Data

AFP case-based surveillance data

### DATA AND RESOURCES

 2012 AFP Case-Based	<span>data valid</span>	<a href="#">Explore</a>
 2013 AFP Case-Based	<span>data valid</span>	<a href="#">Explore</a>
 2014 AFP Case-Based	<span>data valid</span>	<a href="#">Explore</a>
 2015 AFP Case-Based	<span>data valid</span>	<a href="#">Explore</a>
 2016 AFP Case-Based	<span>data valid</span>	<a href="#">Explore</a>
 2017 AFP Case-Based	<span>data valid</span>	<a href="#">Explore</a>

AFP Case-Based

# CKAN Features

## Search & Discovery

- Search all metadata fields
  - ◆ Text search
  - ◆ Map based search
- Facet search
  - ◆ organization, group, format, license, tags
- Power by Apache Solr

The screenshot shows the CKAN Datasets page. At the top left is a sidebar with navigation links: Home, Organizations, Groups, Tags, and Help. The main content area has a breadcrumb trail: Home / Datasets. A green "Add Dataset" button is located at the top right. Below it is a search bar with placeholder text "Search datasets..." and a magnifying glass icon. To the right of the search bar is a dropdown menu labeled "Order by: Last Modified". On the left side, there are three facets: "ORGANIZATIONS" (Archived Dataset (166), AFP (5), Surveillance (4), Admin Data (3), Dashboard (2), Monitoring (2), Guidelines, Manuals... (1)), "GROUPS" (EOC (1), Webinar participants (1)), and "TAGS" (2017 (56), 2016 (33), March (16), 2015 (14), AFP (12), January (12)). The main content area displays 183 datasets found. The first dataset listed is "Webinar datasets", which is marked as PRIVATE and has a CSV download link. The second dataset is "AFP Environmental Sample Dataset", also marked as PRIVATE, with a note about AFP ES data from environmental sample collection sites and processed by the lab, and a CSV download link. The third dataset listed is "AFP Case Verification", marked as PRIVATE, with a note about AFP case verification, and a CSV download link. The fourth dataset listed is "AFP Case-Based Data", marked as PRIVATE, with a note about AFP case-based surveillance data, and a CSV download link.

# CKAN Features

## Data Exploration: Resources Preview

- preview resource data as tables, maps and charts ...
- table previews are interactive; data can be sorted, filtered
- resource with spatial data can be visualized on a map view...

Primary Schools

Manage Go to resource

Education establishment or learning space where children receive primary or elementary education from the age of about five to twelve, coming after preschool and before secondary school

Data Explorer

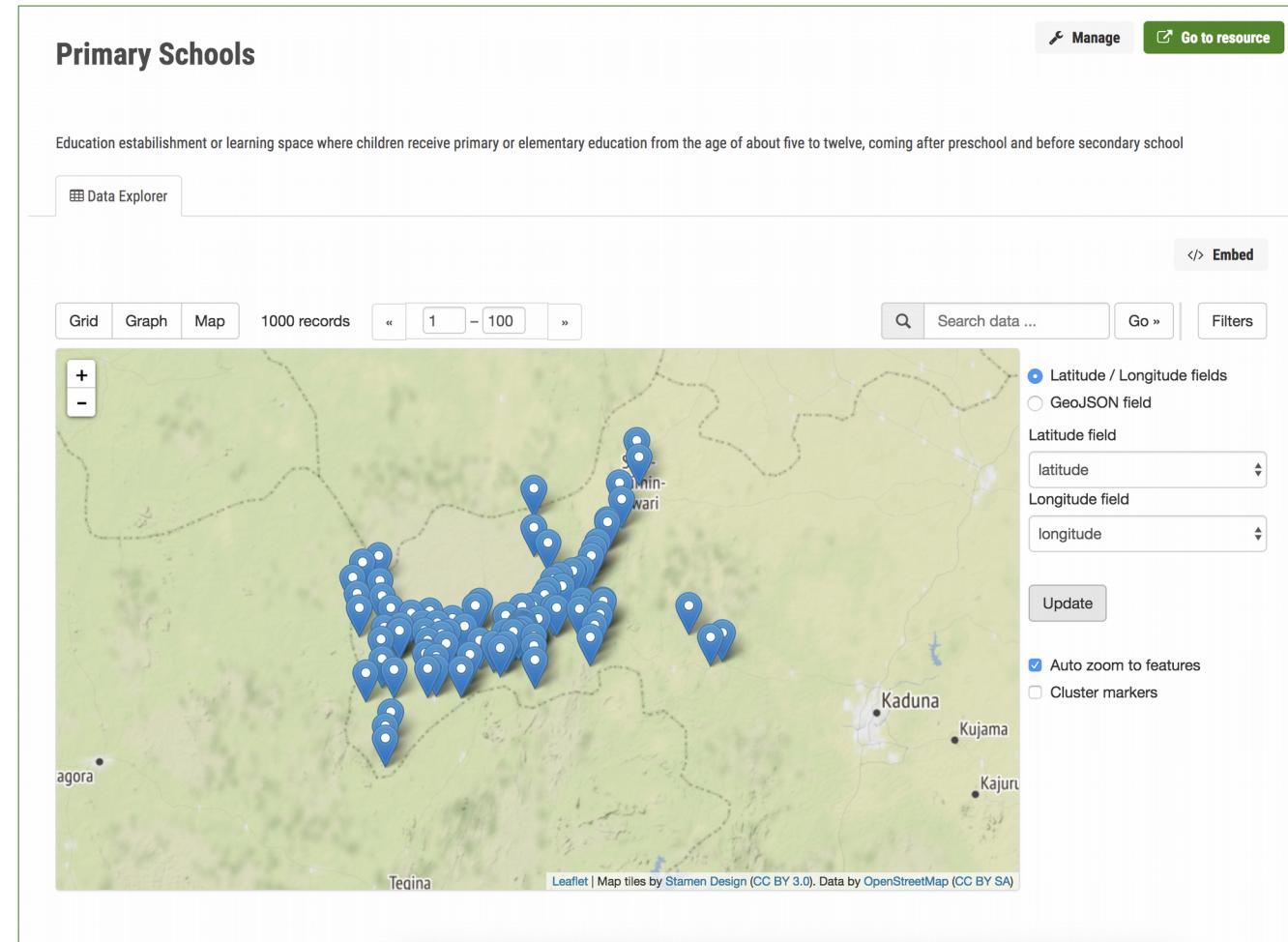
</> Embed

Grid	Graph	Map	1000 records	«	1	– 100	»	Q	Search data ...	Go »	Filters		
FID	geom	latitude	longitude	globalid	name	category	subtype	manage...	education	poi_type	no_of_t...	no_of_st...	wardcode
sv_fc_po...	POINT (6...	10.8068...	6.68934...	f89e087...	Danganji...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.8382...	6.71408...	27d11bc...	Universa...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.8586...	6.72257...	7c63365...	LGEA Pri...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.6367...	6.71253...	723663f...	Universa...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7240...	6.74564...	82f8408...	Alheri Ba...	Primary	Standard	Private	Integrated	School			KD0101
sv_fc_po...	POINT (6...	10.7913...	6.63572...	4ba07af...	LGEA D...	Primary	Standard	Public	Formal	School			KD0101
sv_fc_po...	POINT (7...	10.6354...	7.01701...	aacb34b...	LGEA Sa...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7053...	6.68483...	58d9480...	LGEA Pri...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7098...	6.62616...	716f5e1...	LGEA	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.6665...	6.72380...	b4537ed...	Universa...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7127...	6.61990...	589907a...	Universa...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7134...	6.56642...	0308580...	LEA Pri...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7148...	6.96305...	e710a75...	LGEA Pri...	Primary	Primary	Private	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7277...	6.74585...	2784ceb...	LGEA Pri...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7282...	6.69250...	e4d83ae...	Nomadic...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7411...	6.59851...	88f24b2...	Universa...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7539...	6.60615...	ca48128...	LEA Kun...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (7...	10.6478...	7.04711...	c4056e2...	LGEA Pri...	Primary	Standard	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7623...	6.64210...	1859ac5...	Doka No...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.7781...	6.61835...	69c7867...	Universa...	Primary	Primary	Public	Formal	School			KD0101
sv_fc_po...	POINT (6...	10.6934...	6.73752...	42ca595...	Universa...	Primary	Primary	Public	Formal	School			KD0101

# CKAN Features

## Data Visualization: Map Preview

- preview resource data as tables, maps and charts ...
- table previews are interactive; data can be sorted, filtered
- resource with spatial data can be visualized on a map view...



## CKAN Features

### FileStore & DataStore:

- CKAN manages uploaded files through the FileStore
- CKAN stores structured data from resources into the DataStore (*ad hoc database*) and makes content searchable
- Implemented as extensions



# CKAN Features

## Extensible:

- some core functionalities are implemented as extensions
- easily customize, disable or swap out existing features
- CKAN community provides many more extensions that extend the capabilities of CKAN

### Extension listing

<https://extensions.ckan.org>

<https://ckan.org/category/extensions/>

<https://ckan.viderum.com/extensions/>



VIDERUM

CKAN Cloud ▾ CKAN Services Blog About Get Started

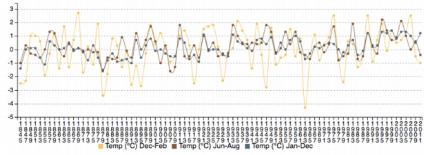
## Advanced map views

Add thematic (choropleth) maps to CKAN. You can display e.g. the population density of various regions by colouring the map in different shades of blue. Also regions/countries can be clickable and display more data after being selected. Works with GeoJSON files.



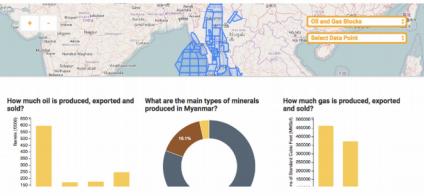
## Advanced data visualisations

Allow users to visualise your datasets using graphs, and charts. You can also pre-define and save visualisations for selected datasets to provide users with a better understanding of your data. An embed function provides code snippets so charts can be embedded within external websites.



## Dashboard

Our data dashboard visualises selected datasets on a single overview screen. You can add graphs, charts, and maps to give users a quick summary of the data you value most. Demonstrate your organisations growth and visualise the effects of your policies and programs. The dashboard is simple to use, self-explanatory and includes all [Read more](#) 



## Request data

Gather feedback from your community and let users request datasets they would like to see. The extension provides a form to make requests and allows users to follow up with the publishing process.

**Suggest a Dataset**

Suggest a Dataset enables users to recommend datasets not yet published in the organization. If you want a dataset and you are not able to find it among the published datasets, you can let us know via suggestion. Please provide as much information as you can, you can also specify an organization responsible for

**Suggest a dataset**

\* Title: eg. Dataset Name

\* Description: Data Description

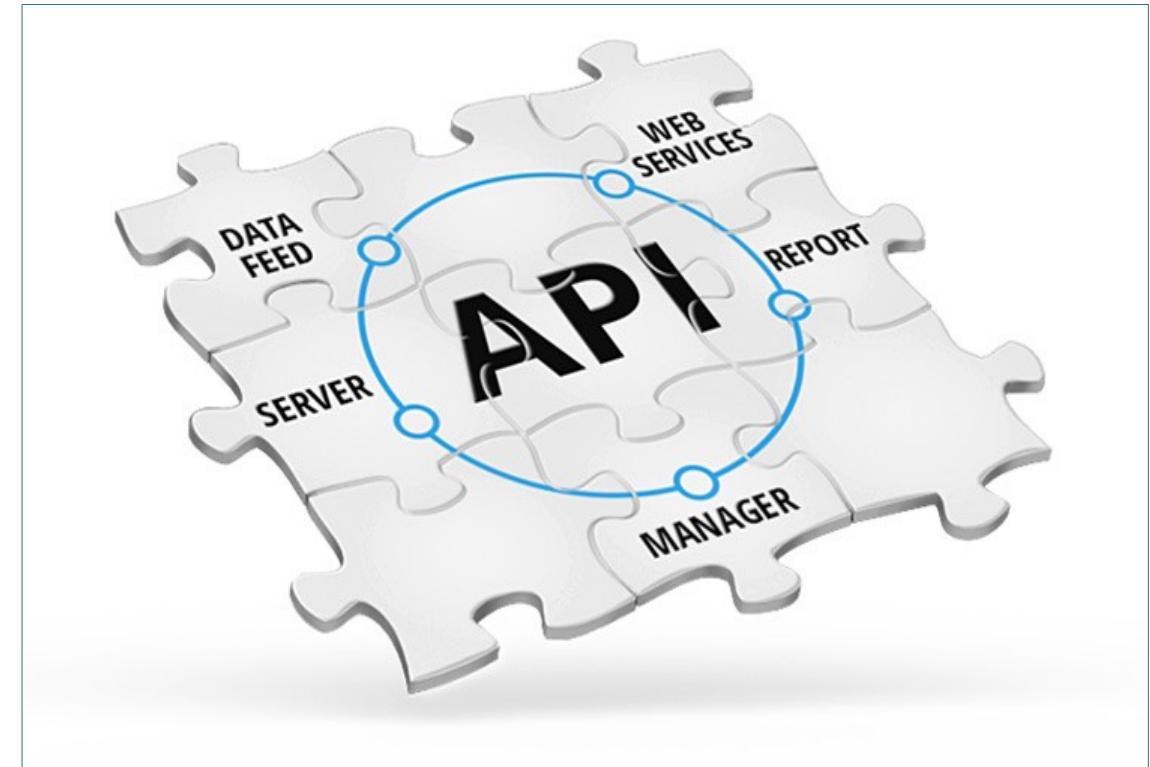
You can use Markdown formatting here

Suggested

## CKAN Features

### Rich API for Data Access:

- all operations possible via the web interface can equally be performed via the API
- token based access
- RPC Style
- consumes and returns JSON data



# Publishing Data on CKAN



# Data Organization

## Organizations

- cataloging structure
- an organization collects together datasets published and managed by a single entity
- equally defines the users with authorization to create, edit and view (private) datasets

The screenshot shows the 'Nigeria Polio EOC' organization page on a data portal. The top navigation bar includes links for Home, Datasets, Organizations, Groups, Charts, and About, along with a search bar. The main content area displays the organization's logo, name, and a brief description: 'A data portal for the Nigeria Polio Emergency Operation Center'. It shows 0 followers and 166 datasets. A 'Follow' button is present. Below this, sections for Organizations (Nigeria Polio EOC), Groups (EOC 1), and Tags (2017 (55), 2016 (33)) are listed. To the right, a sidebar for 'Nigeria Polio EOC sub-organizations' includes a search bar, an 'Include Sub-Organizations' checkbox, and an 'Order by: Relevance' dropdown. Three dataset cards are visible: 'Week 30 AFP CB data' (private, application/msaccess), 'Wk25 2017 ES Data' (private, application/msaccess), and 'Wk25 AFP Lab-based data' (private, application/msaccess).

# Data Organization

## Groups

- another cataloging structure
- collects together similar or related datasets found across organizations
- equally useful for managing authorization to subsets of datasets within an organization

The screenshot shows a dataset group page for the "Health" group on the Polio Emergency Operations Center website. The top navigation bar includes links for Home, Datasets, Organizations, Groups, Charts, and About. A search bar is located in the top right corner.

The main content area displays the "Health" group details, including its icon (a caduceus), follower count (0), dataset count (1), and a "Follow" button. Below this, sections for Organizations (Nigeria Polio EOC), Groups (Health), and Tags (2017, March) are shown. A search bar allows users to search for datasets.

A single dataset, "March NIPDs 2017 Tally Sheet Data", is listed. It is marked as PRIVATE and contains aggregate vaccination data. The file is available in XLSX format.

Filtering options include "Datasets", "Activity Stream", and "About". The results are ordered by Relevance.

# Datasets & Resources

**DATASET** collect together similar or related data

- has one or more resources
- owned by a single organization
- can belong to one or more groups
- has a single
  - ◆ source
  - ◆ license
  - ◆ author
  - ◆ maintainer
  - ◆ visibility
  - ◆ ...

**RESOURCE** can be uploaded data files, url links to external resources or API endpoints

- represents actual data being published
- has one format
- those grouped under the same dataset represent
  - same data in different formats
  - same data but different time periods

# Creating Datasets

- A two-step process with last step being creating at least 1 resource for the dataset
- Form provides predefined metadata fields for datasets
- Form has provision for custom fields to collect extra metadata defined by user
- Some metadata fields:
  - Title
  - Organization
  - Visibility
  - License
  - ...

1 Create dataset

2 Add data

**Title:** eg. A descriptive title

\* URL: [data.eocng.org/dataset/<dataset>](http://data.eocng.org/dataset/<dataset>) Edit

**Description:** eg. Some useful notes about the data

You can use [Markdown formatting here](#)

**Tags:** eg. economy, mental health, government

**License:** License not specified

ⓘ License definitions and additional information can be found at <http://opendefinition.org/>

**\* Organization:** AFP

**Visibility:** Private

**Source:** <http://example.com/dataset.json>

**Version:** 1.0

**Author:** Joe Bloggs

**Author Email:** joe@example.com

**Maintainer:** Joe Bloggs

**Maintainer Email:** joe@example.com

The *data license* you select above only applies to the contents of any resource files that you add to this dataset. By submitting this form, you agree to release the *metadata* values that you enter into the form under the Open Database License.

\* Required field

**Next: Add Data**

# Creating Resources

- Upload data file or provide link to external resource
- Form fields:
  - Name
  - Description
  - Format
  - ...

1 Create dataset

2 Add data

**Title:** eg. A descriptive title

\* URL: [data.eocng.org/dataset/<dataset>](http://data.eocng.org/dataset/<dataset>) Edit

**Description:** eg. Some useful notes about the data

You can use [Markdown formatting here](#)

**Tags:** eg. economy, mental health, government

**License:** License not specified

ⓘ License definitions and additional information can be found at <http://opendefinition.org/>

**\* Organization:** AFP

**Visibility:** Private

**Source:** <http://example.com/dataset.json>

**Version:** 1.0

**Author:** Joe Bloggs

**Author Email:** joe@example.com

**Maintainer:** Joe Bloggs

**Maintainer Email:** joe@example.com

The *data license* you select above only applies to the contents of any resource files that you add to this dataset. By submitting this form, you agree to release the *metadata* values that you enter into the form under the Open Database License.

\* Required field

Next: Add Data

# Create using CKANAPI

A command line interface and Python module  
for accessing the CKAN Action API.

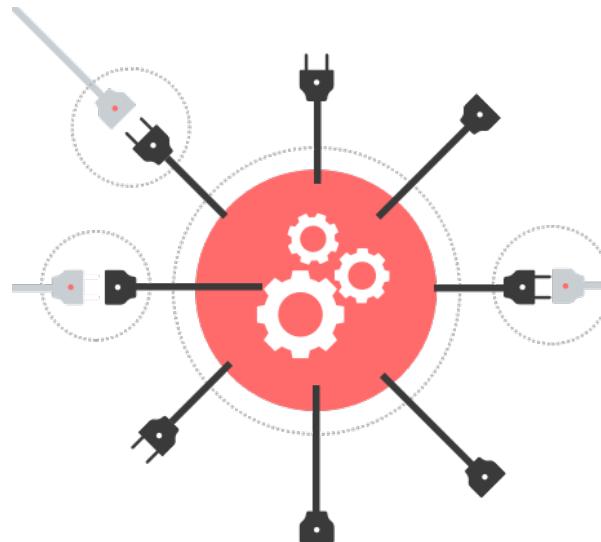
<https://github.com/ckan/ckanapi>

Supported operations:

- Bulk loading or dumping of data
- Bulk deletion
- Shell pipelines

Target objects:

- organizations & groups
- datasets and resources
- users



# Managing Organization Users

- users granted organization membership can have any of these roles: *admin, editor, member*
- users with admin or editor roles can create datasets; users with member role can only view organization datasets
- users with admin role can invite existing or new users to their organization

The screenshot shows a user interface for adding a member to an organization. At the top, there are navigation links: 'Edit' (with a pencil icon), 'Members' (with a paw print icon), and 'View' (with an eye icon). Below these, the title 'Add Member' is centered. To the right of the title is a link 'Back to all members' with a left arrow icon.

**Existing User :**  
If you wish to add an existing user, search for their username below.  
 ▼

**New User :**  
If you wish to invite a new user, enter their email address.

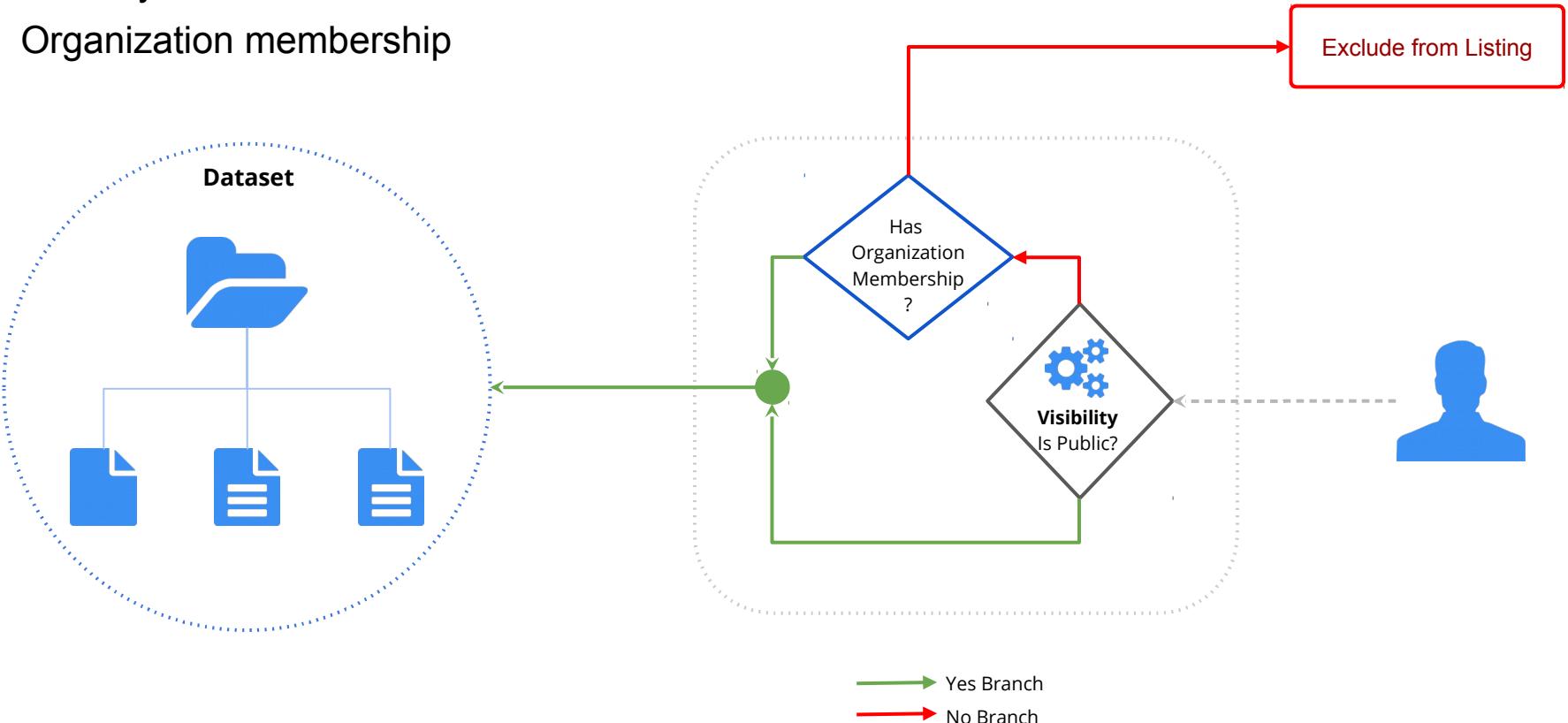
**OR**

**Role:**  
 ▼

**Add Member**

# Access Control

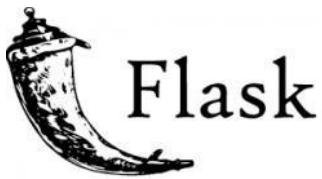
- Restrict access to datasets with sensitive information
- Visibility: **Public** vs **Private**
- Organization membership



# CKAN Architecture



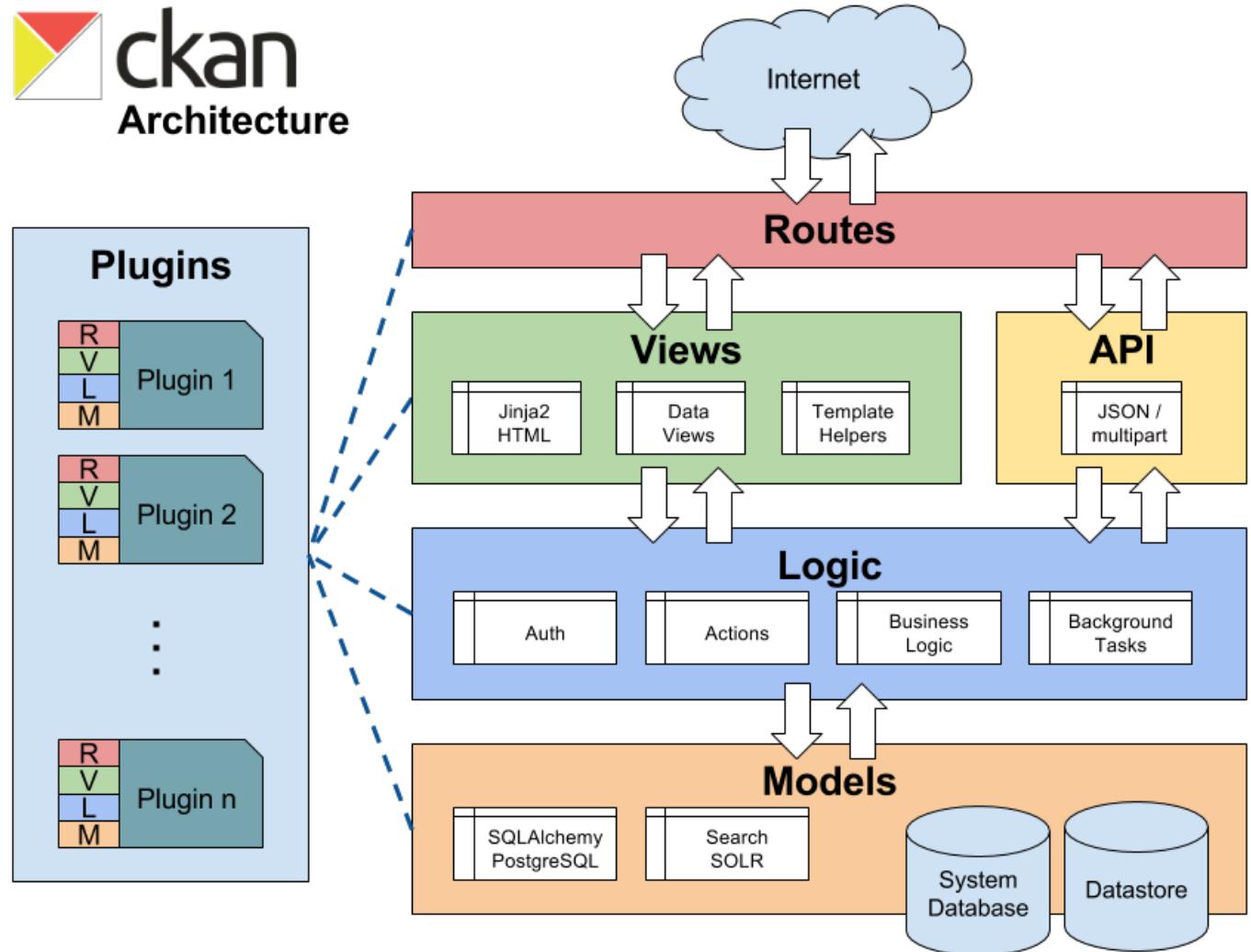
# Apps, Frameworks, Library & Tools





# Code Architecture

- Plugin-based
- Server rendered templates
- MVC design pattern
- RPC Styled API



# CKAN Experience

Experience  
deserving, encouraging, skill

# Geo-Referenced Infrastructure and Demographic Data for Development (GRID3)

<https://grid-nigeria.org/>

- customized cataloguing structures:
- organizations → states
  - groups → sectors
- customized data access model
- extended map search
- transform resources to various geospatial forms
- repurposed cataloguing implementation for the help center
- ...

The screenshot shows the GRID3 Nigeria website. At the top, there is a navigation bar with links for Home, Datasets, States, Sectors, About, Help Center, and a search bar. The main header features the GRID3 logo and the text "GEO-REFERENCED INFRASTRUCTURE AND DEMOGRAPHIC DATA FOR DEVELOPMENT (GRID3) PROJECT". Below the header is a sub-header: "Generate integrated mapping of development needs throughout Nigeria across the health, education, environment, agriculture, urban planning and investment promotion sectors." The page includes a map of Nigeria with state boundaries and a zoom control (+/-). To the right, there are six categories with icons: Agriculture (green plants), Business and Economy (person with chart), Communication (tv), Crime and Justice (scales of justice), Education (graduation cap), and Emergency Services (ambulance). At the bottom, there are sections for Related Links (About Portal, Terms of use, Privacy Policy), Newsletter (FAQs, Contact Us), and Partners (Federal Republic of Nigeria, Bill & Melinda Gates Foundation, DFID Nigeria, eHealth Africa).

# Polio Emergency Operations Center (EOC)

<https://data.eocng.org/>

- Resource validation
- Gather2 integration
- Sub-Organizations

Log in Newsletter FAQ

Search 

Polio Emergency Operations Center

Home Charts About

Count up from last polio report  
751 : 15 : 17 : 03  
DAYS HOURS MINUTES SECONDS



“ At this last phase of polio eradication in Nigeria, only a quality, timely and complete data generated from field, analyzed and transformed into meaningful report/result can guarantee us better understanding of actual and real situation for informed decision making.

Dr. Faisal Shuaib | ED, NHPGDA

**Immunization**  
Immunization coverage data available at LGA, State and National levels to aid identification and addressing of challenges, and monitoring of trends.

**Surveillance**  
Data from monitoring and assessing the impact of strategies and activities for reducing morbidity and mortality of vaccine-preventable diseases.

**GIS**  
Data Portal provides public access to all of our published GIS data that can be used for visualizing and making location more accessible.

1 Dataset    6 Organizations    4 Groups    1 Resource

Most recent datasets  
16 08 2017 EOC Data Management Portal ... Non-administrative User Guide intended to give assistance to people that will be using the portal

**Partners**

World Health Organization  
unicef  
CDC  
coregroup  
Rotary  
BILL & MELINDA GATES foundation

© 2018 Polio Emergency Operations Center. All rights reserved except published datasets/resources and metadata.  
All datasets/resources including metadata published on data.eocng.org are licensed under Polio Emergency Operations center

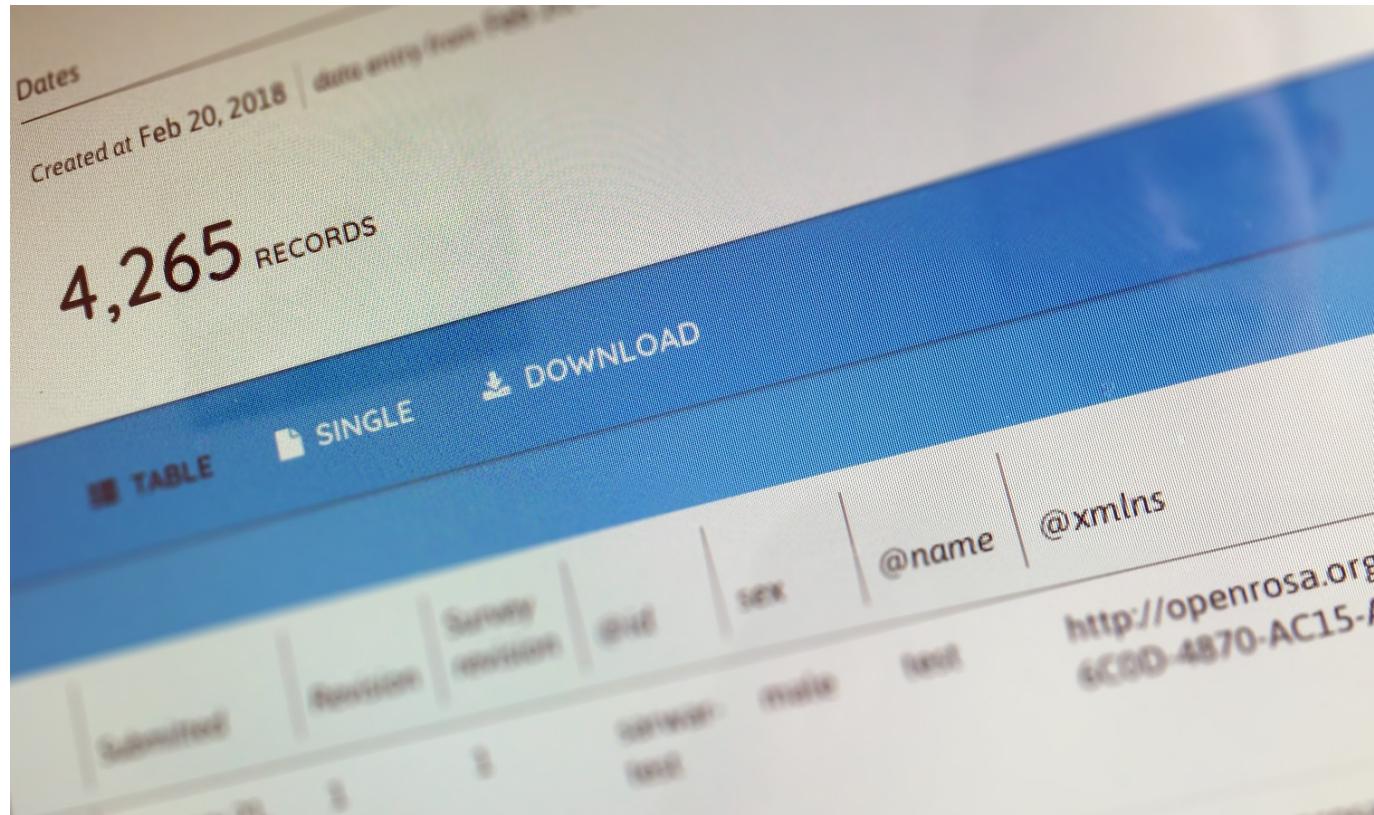
About Portal Terms of use Newsletter FAQs Contact Us



## An ODK-Compatible Data Collection Platform

- versatile, secure and performant platform for data collection
- developed by eHealth Africa
- built using open, highly scalable technologies
- built on top of Aether



# Aether Platform



## Aether...

...is a reliable and secure development platform created by eHealth Africa to enable organizations to build solutions that curate and exchange live information.

It utilizes “data contracts” between systems, simplifies the movement of data between applications and helps developers adhere to best practices for e-health systems design.



A large, stylized text graphic reading "Thank You". The letters are white with a thick blue outline. The entire graphic is suspended by a thin brown string from a small metal hook at the top center.

- [abdulhakeem.shaibu@ehealthnigeria.org](mailto:abdulhakeem.shaibu@ehealthnigeria.org)
- [hkmshb@gmail.com](mailto:hkmshb@gmail.com)