

# CKAN

# Tech Roadmap [Thoughts]

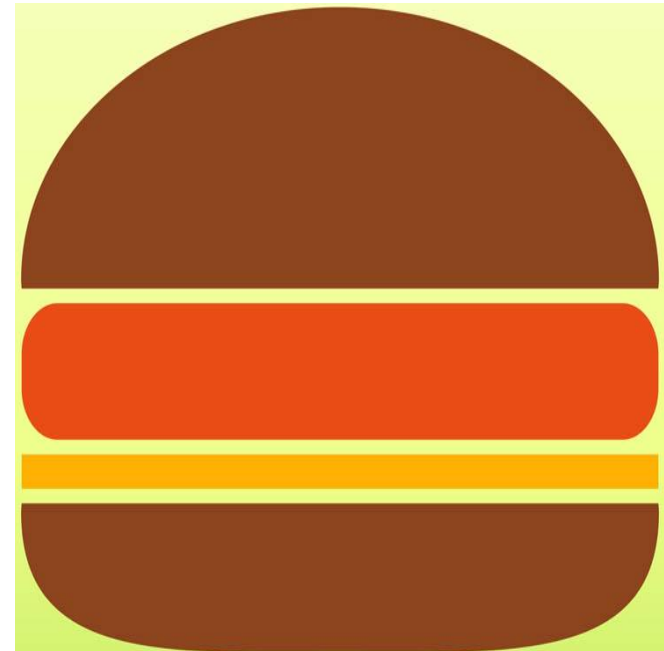
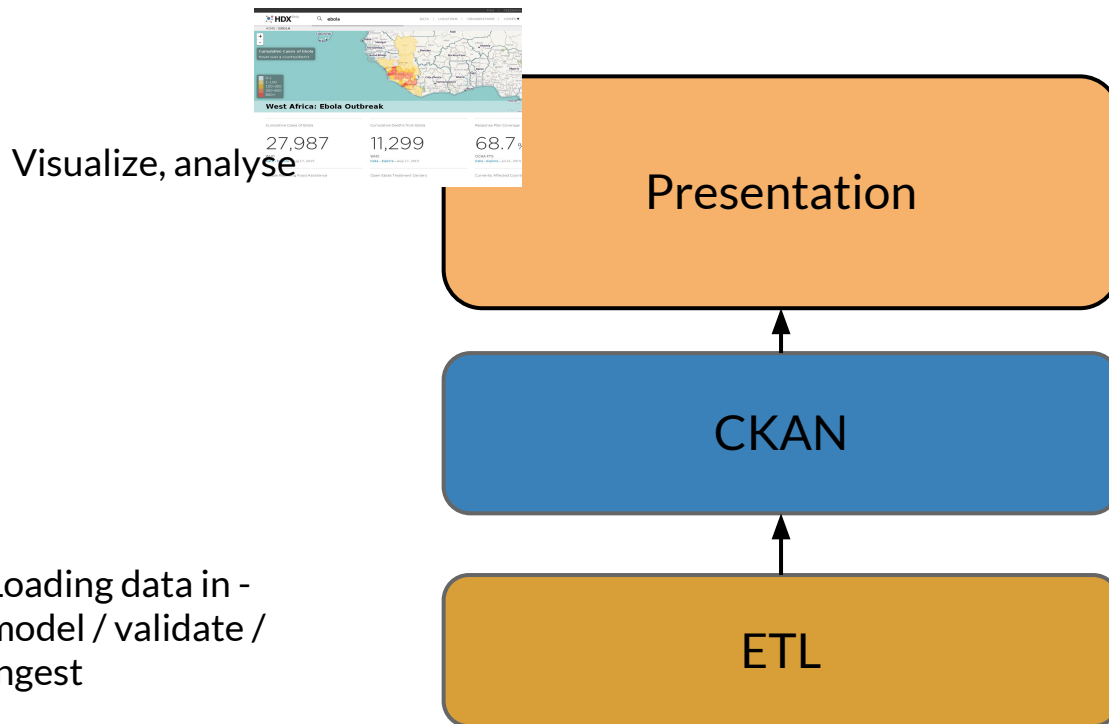
PRESENTED BY

**Rufus Pollock**



# Hamburger Analogy

- CKAN = meat in the hamburger.
- Presentation & ETL = “bun” around the burger
  - modules outside of “core” code - interacting via (web) API.
- Separation in tech architecture does **not** require they be separate experience for users -- the experience can be seamless (e.g. they could be built in JS + HTML and embedded into your CKAN webspace).
- But separating architecturally and separately makes it easier and faster to develop.



# What's Working

- Very solid and mature platform
- Large number of deployments including many major, long-running ones
  - Respected backwards compatibility
- Metadata workflow very solid
- FileStore and DataStore now multi-year developments with substantial deployments

# Not working as well as we want

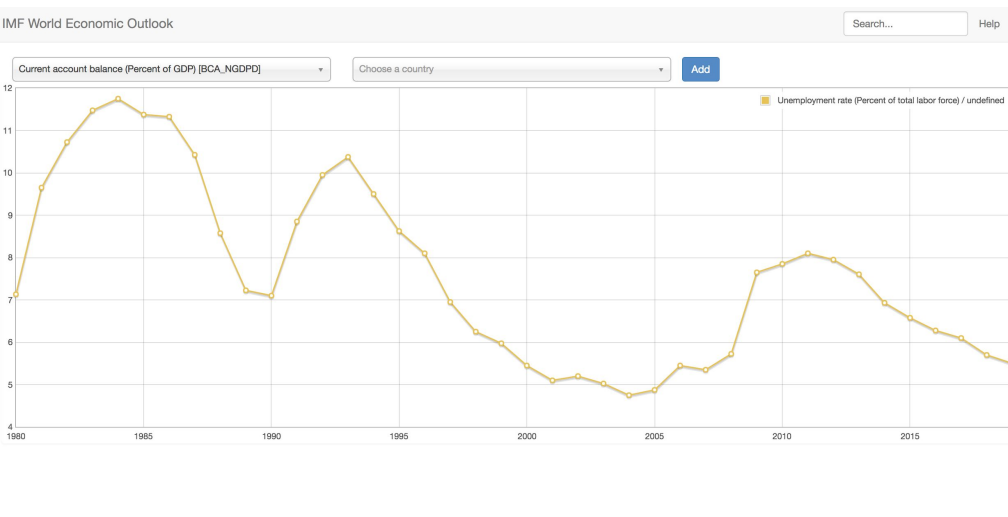
- Legacy: e.g. Pylons vs Flask, sqlalchemy-migrate, vdm
- Presentation
  - View creation and saving - no way for visiting user to create and save views
  - Dashboards
  - Rich Data APIs
- Data import - want to allow rich, powerful ETL integration
- Datastore (?)

# How Should Views Work

- Anyone (logged in) can create and save
- Wizard for creation
- Create using data from multiple datasets
- Massively improved version of (very crude)

<http://dev.rufuspollock.org/ckan-explorer/?endpoint=https://old.datahub.io>

- <http://dev.rufuspollock.org/imf-weo/>



CKAN Data Explorer

SQL Query

Query the data using SQL via the database (SQL)

```
SELECT * FROM "WORLD" WHERE "Year" > 2000
```

Query

Year	Value	Unit	Source
2000	7.5	%	World Bank
2001	7.5	%	World Bank
2002	7.5	%	World Bank
2003	7.5	%	World Bank
2004	7.5	%	World Bank
2005	7.5	%	World Bank
2006	7.5	%	World Bank
2007	7.5	%	World Bank
2008	7.5	%	World Bank
2009	7.5	%	World Bank
2010	7.5	%	World Bank
2011	7.5	%	World Bank
2012	7.5	%	World Bank
2013	7.5	%	World Bank
2014	7.5	%	World Bank
2015	7.5	%	World Bank

Healthy Life Expectancy at birth (female/male average) in Netherlands 2009 - 2014

Healthy Life Expectancy at birth (female/male average) in Netherlands 2009 - 2014

Healthy Life Expectancy at birth (female/male average) in Netherlands 2009 - 2014

Healthy Life Expectancy at birth (female/male average) in Netherlands 2009 - 2014

# How should ETL work

- Analyse - guess types
- Model
- Data validate
- Ingest

# Data Packages

- Key feature of intermediate “(meta)-data bus” layer (CKAN) is providing standard set of metadata and way to access it.
- Adopting Data Packages would build on and improve what is currently there
- Key metadata:
  - a. Dataset and Resource (file) – name, title, description etc
  - b. Data structure (syntax) – fields, types etc
  - c. Quality information (e.g. validation errors, size, up-to-dateness)
- In current CKAN: (a) is well-covered, (b) is semi-covered by current DataStore implementation and (c) has been discussed but never formalized
- Data Package standardizes (a), Table Schema does (b) for tabular data, GoodTables has a start on (c) but much to do