Bootstrapping the open data ecosystem

Dr. Rufus Pollock

Fellow in Economics
University of Cambridge

Director
Open Knowledge Foundation

UNIVERSITY OF CAMBRIDGE

Open Knowledge Foundation

We are living in a digital age.

Technological Explosion

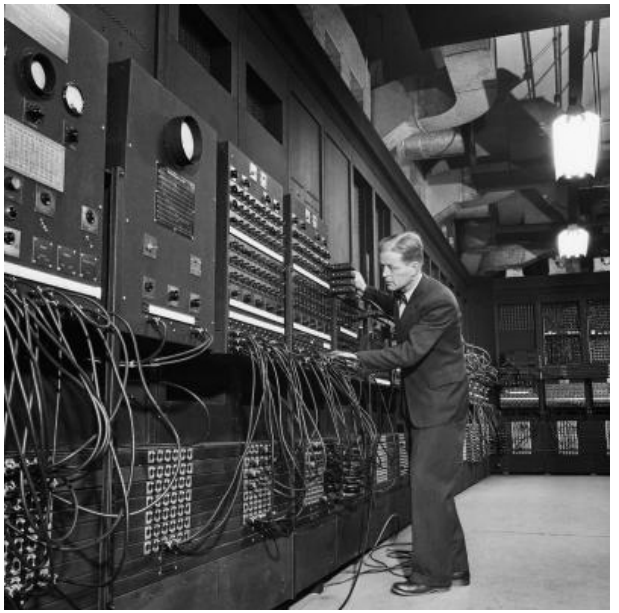
Powerful computers, cheap storage
global communication networks

A shared ecosystem of data?

Social and Commercial Benefits

Life-saving drugs,
improved public services,
increased access to cultural and educational materials ...

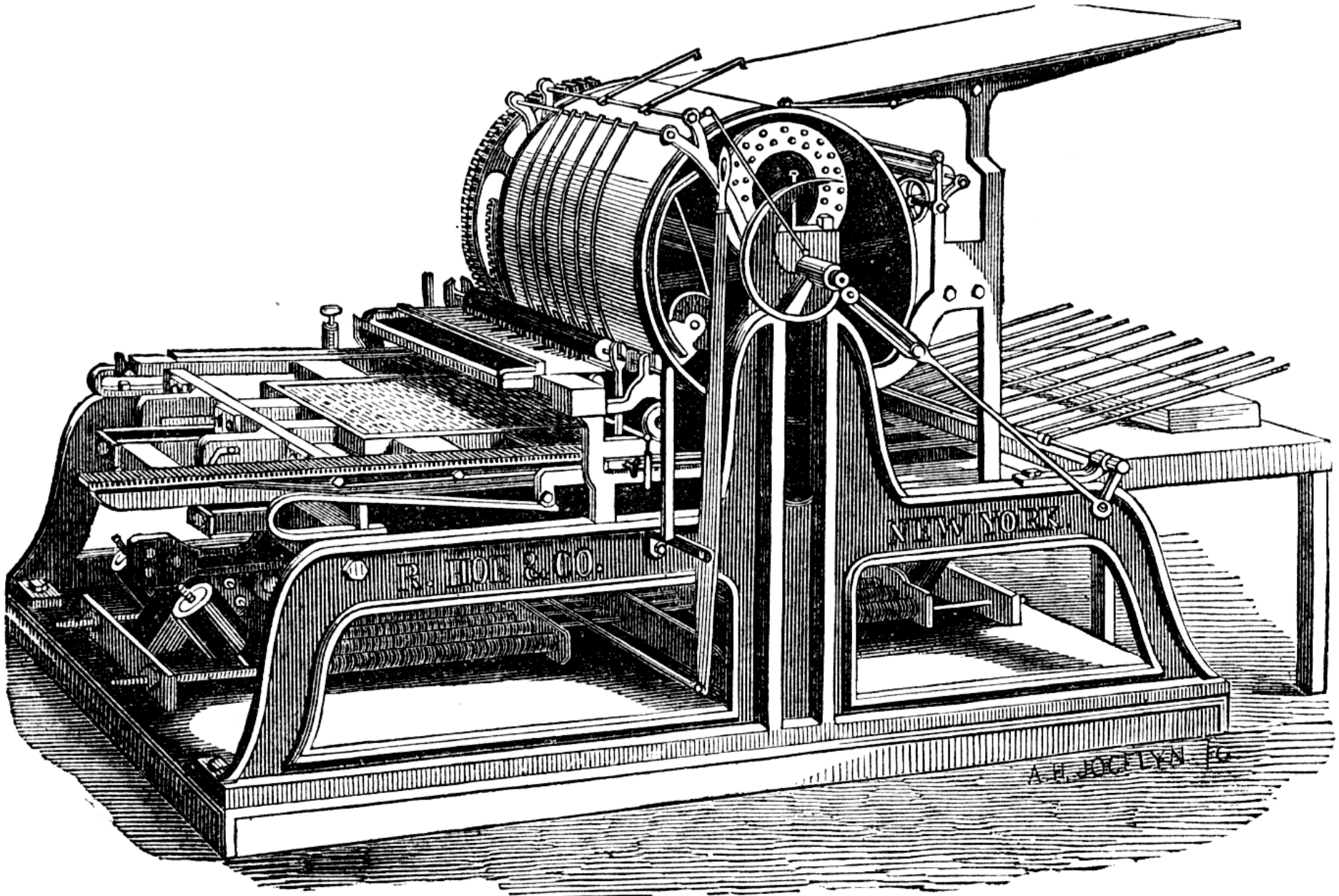
From books to bits…



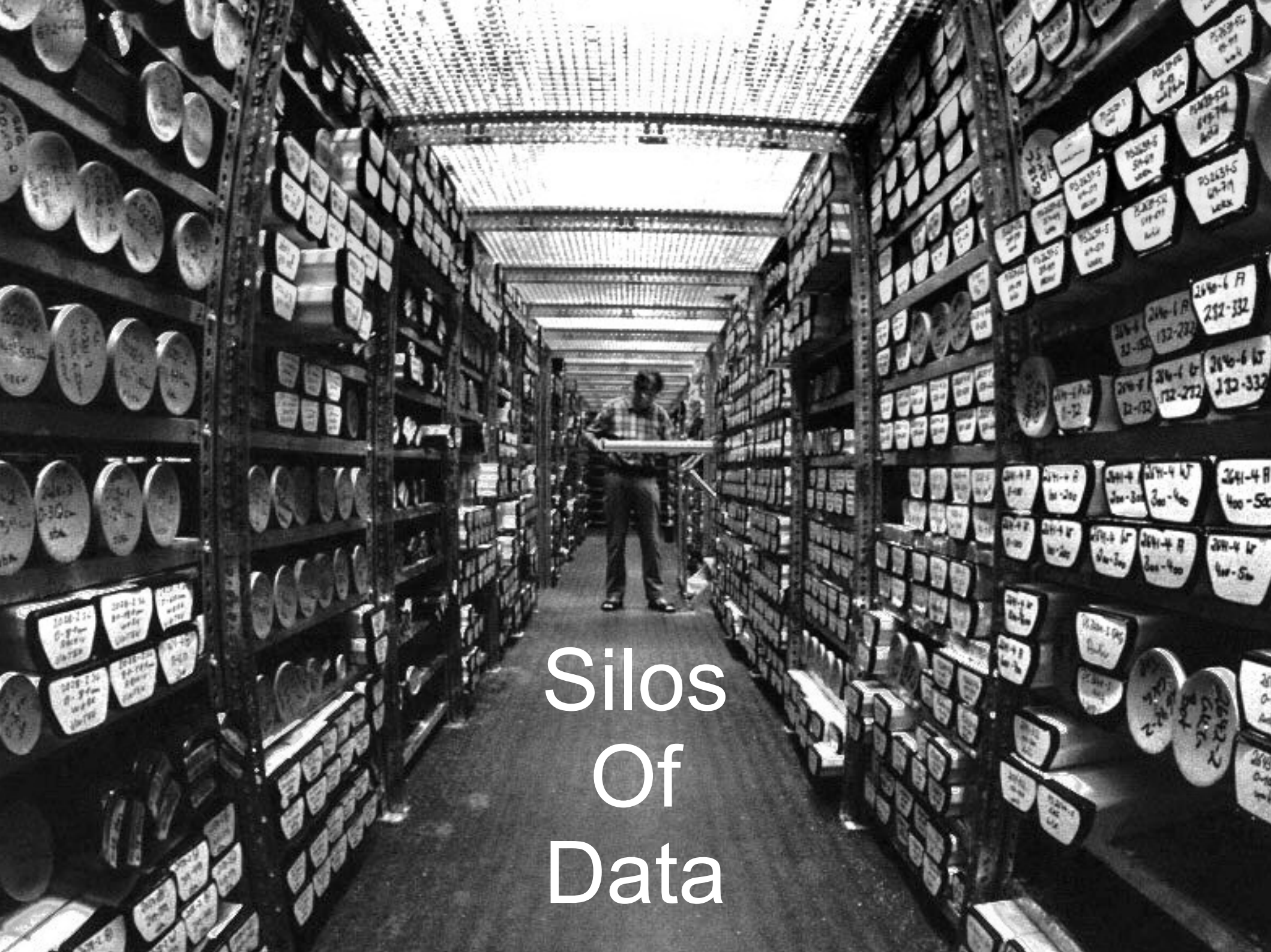
But We're Still Working Out the Details!

ONLINE
≠
OPEN
lenovo
T500

Copyright and IP Law



HOE'S ONE-CYLINDER PRINTING PRESS.


Silos
Of
Data

open knowledge

=

any content or data that anyone is free to use, reuse or redistribute

All forms of content and data

From sonnets to statistics,
genes to geodata…

Some familiar names...

Wikipedia / Wikimedia Projects
Open Street Map
Flickr Commons
Freebase
Geonames
Dbpedia
Project Gutenberg / Runeberg
Librivox
MusicBrainz
Internet Archive / Open Library
Connexions
... and many many more!

Confusing Signals

aldridge
traffic
systems

The best thing to do
with your data will
be thought of by
someone else!
OPEN DATA
DATA
DATA
DATA
DATA

Technically Open

(As well as legally open)

Shiny PDFs and Shiny Front-Ends Aren't What We Want

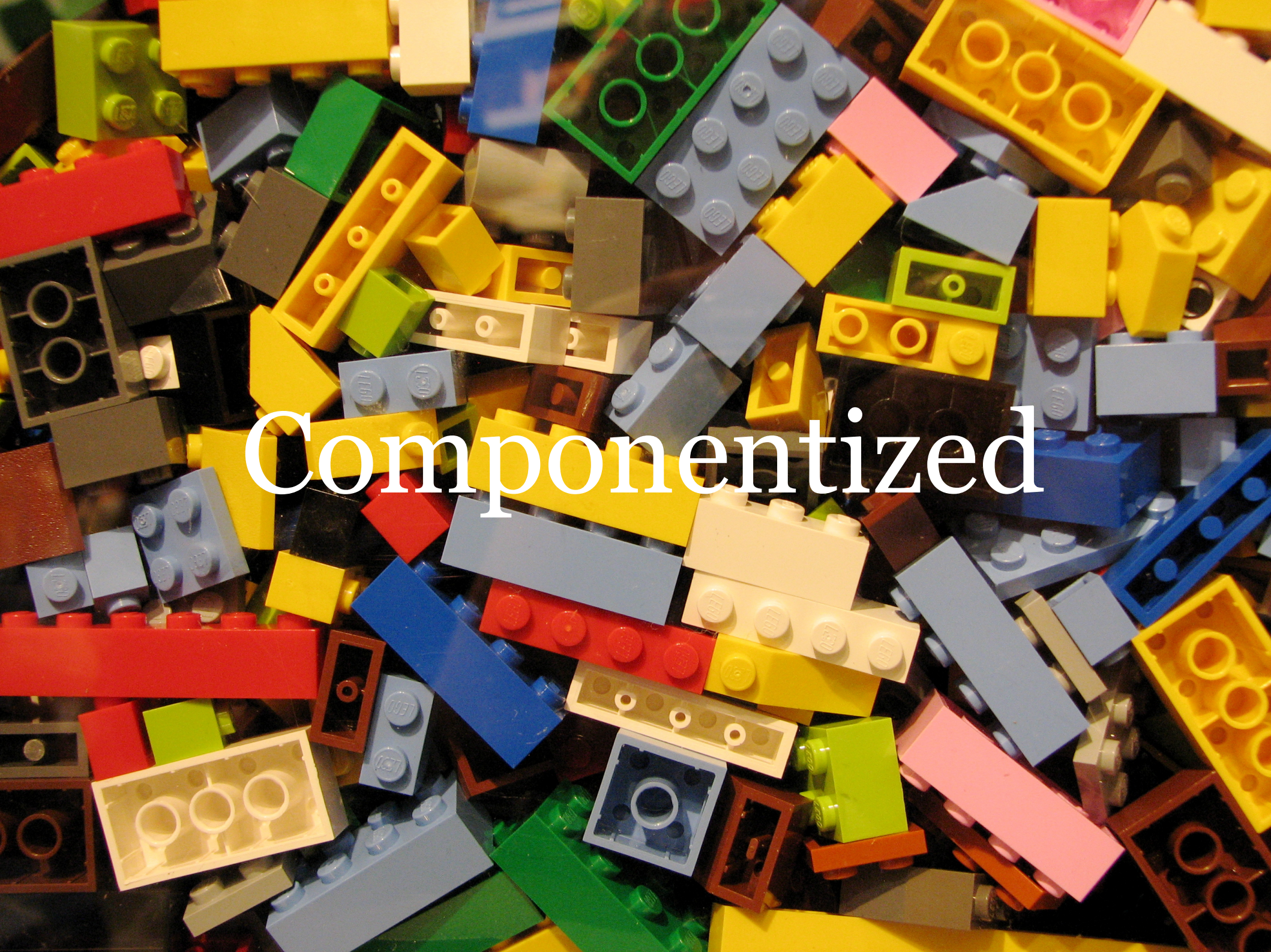
“Raw Data Now!”

How Do We Scale?
What Approach is Required
for Large Scale Collaboration
On Data and Content
“Development”?

Borrow from Free/Open Source

Develop a model that is:

Incremental
Collaborative
Decentralised
Componentised


Componentized

We Can Only Handle So Much Complexity

Create and Share Data Packages

Developing a“Debian” of Data

18,000 Packages
Maintained by 100s of people

The same is possible for data

Componentization allows us to scale
Openness allows us to put
Humpty-Dumpty back together again

Community
Driven

To Many Eyes All Bugs Are Shallow

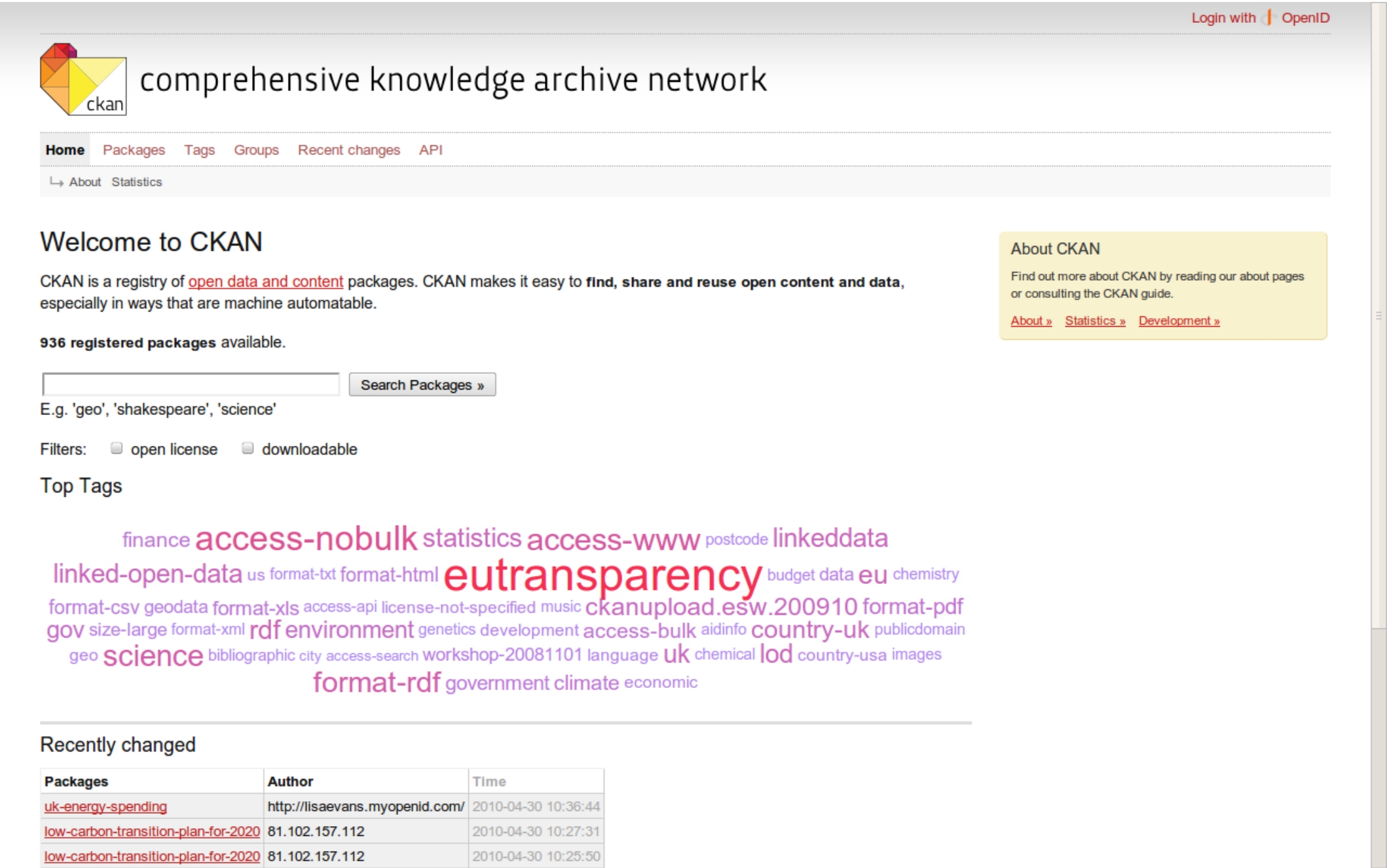
Collaborative developments
Can yield large benefits

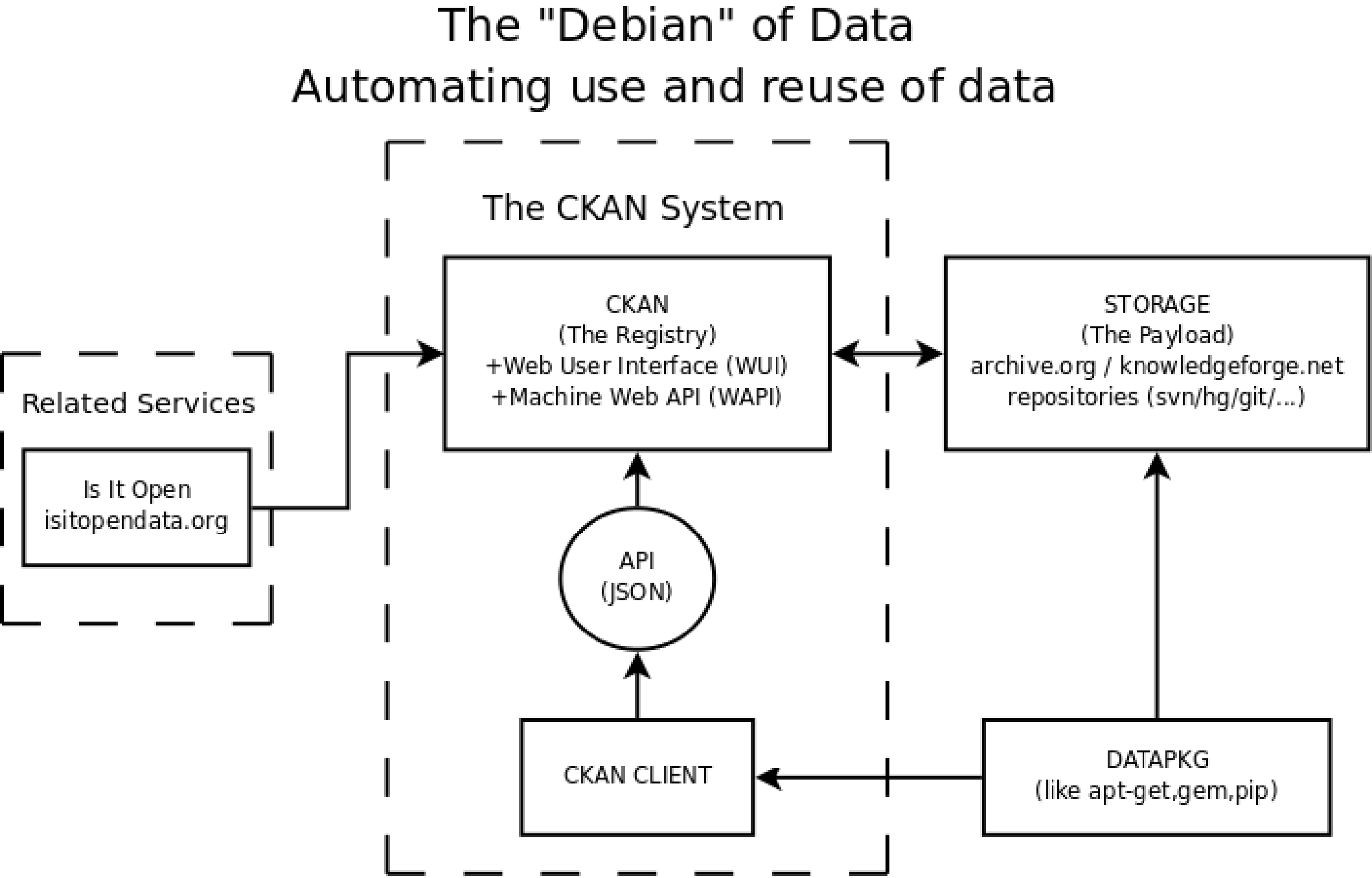
OSM and Haiti
MP Expenses

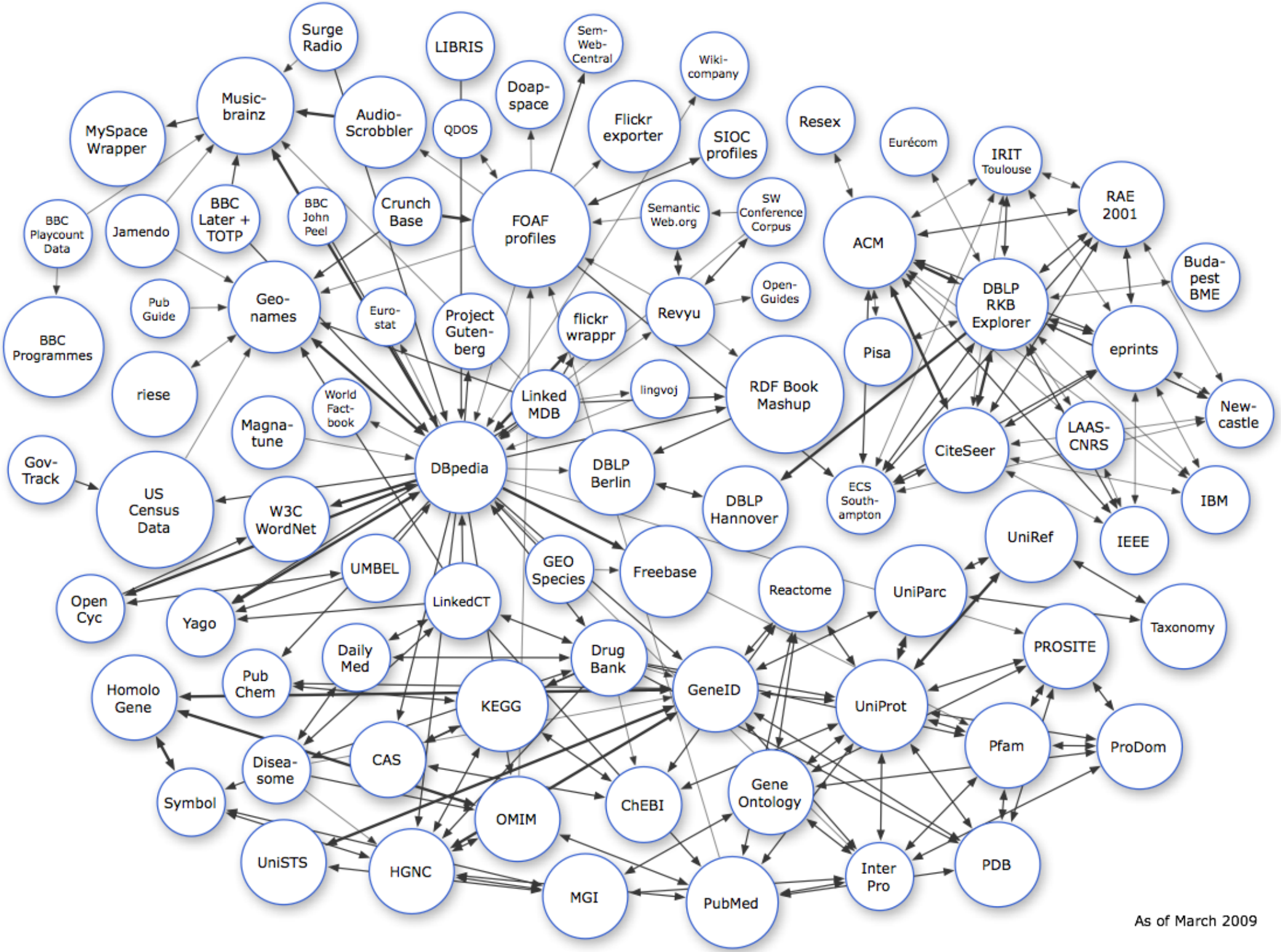
Postboxes to Galaxies, Biodiversity to Dictionaries

Developing the Tools

CKAN


Login with OpenID
ckan
comprehensive knowledge archive network
Home Packages Tags Groups Recent changes API
About Statistics
Welcome to CKAN
CKAN is a registry of open data and content packages. CKAN makes it easy to find, share and reuse open content and data, especially in ways that are machine automatable.
936 registered packages available.
Search Packages »
E.g. 'geo', 'shakespeare', 'science'
Filters: open license downloadable
Top Tags
finance access-nobulk statistics access-www postcode linkeddata
linked-open-data us format-txt format-html eutransparency budget data eu chemistry
format-csv geodata format-xls access-api license-not-specified music ckanupload.esw.200910 format-pdf
gov size-large format-xml rdf environment genetics development access-bulk aidinfo country-uk publicdomain
geo science bibliographic city access-search workshop-20081101 language uk chemical lod country-usa images
format-rdf government climate economic
About CKAN
Find out more about CKAN by reading our about pages or consulting the CKAN guide.
About » Statistics » Development »
Recently changed
Packages Author Time
uk-energy-spending http://lisaevans.myopenid.com/ 2010-04-30 10:36:44
low-carbon-transition-plan-for-2020 81.102.157.112 2010-04-30 10:27:31
low-carbon-transition-plan-for-2020 81.102.157.112 2010-04-30 10:25:50


The "Debian" of Data
Automating use and reuse of data
The CKAN System
CKAN
(The Registry)
+Web User Interface (WUI)
+Machine Web API (WAPI)
STORAGE
(The Payload)
archive.org / knowledgeforge.net
repositories (svn/hg/git/...)
Related Services
Is It Open
isitopendata.org
API
(JSON)
CKAN CLIENT
DATAPKG
(like apt-get,gem,pip)


Surge Radio
LIBRIS
Sem-Web-Central
Wiki-company
Music-brainz
Audio-Scrobbler
Doap-space
Flickr exporter
SIOC profiles
Resex
Eurécom
IRIT Toulouse
MySpace Wrapper
QDOS
RAE 2001
BBC Playcount Data
Jamendo
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
Semantic Web.org
SW Conference Corpus
ACM
Buda-pest BME
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Revyu
Open-Guides
DBLP RKB Explorer
BBC Programmes
Pisa
eprints
riese
World Fact-book
Linked MDB
lingvoj
RDF Book Mashup
New-castle
Magna-tune
Gov-Track
DBpedia
DBLP Berlin
CiteSeer
LAAS-CNRS
US Census Data
W3C WordNet
DBLP Hannover
ECS South-ampton
IBM
UMBEL
GEO Species
Freebase
UniRef
IEEE
Open Cyc
Yago
LinkedCT
Reactome
UniParc
Taxonomy
Daily Med
Drug Bank
PROSITE
Homolo Gene
Pub Chem
KEGG
GeneID
UniProt
Pfam
ProDom
Disea-some
CAS
Gene Ontology
Symbol
ChEBI
OMIM
UniSTS
HGNC
MGI
PubMed
Inter Pro
PDB
As of March 2009

PROTOTYPE

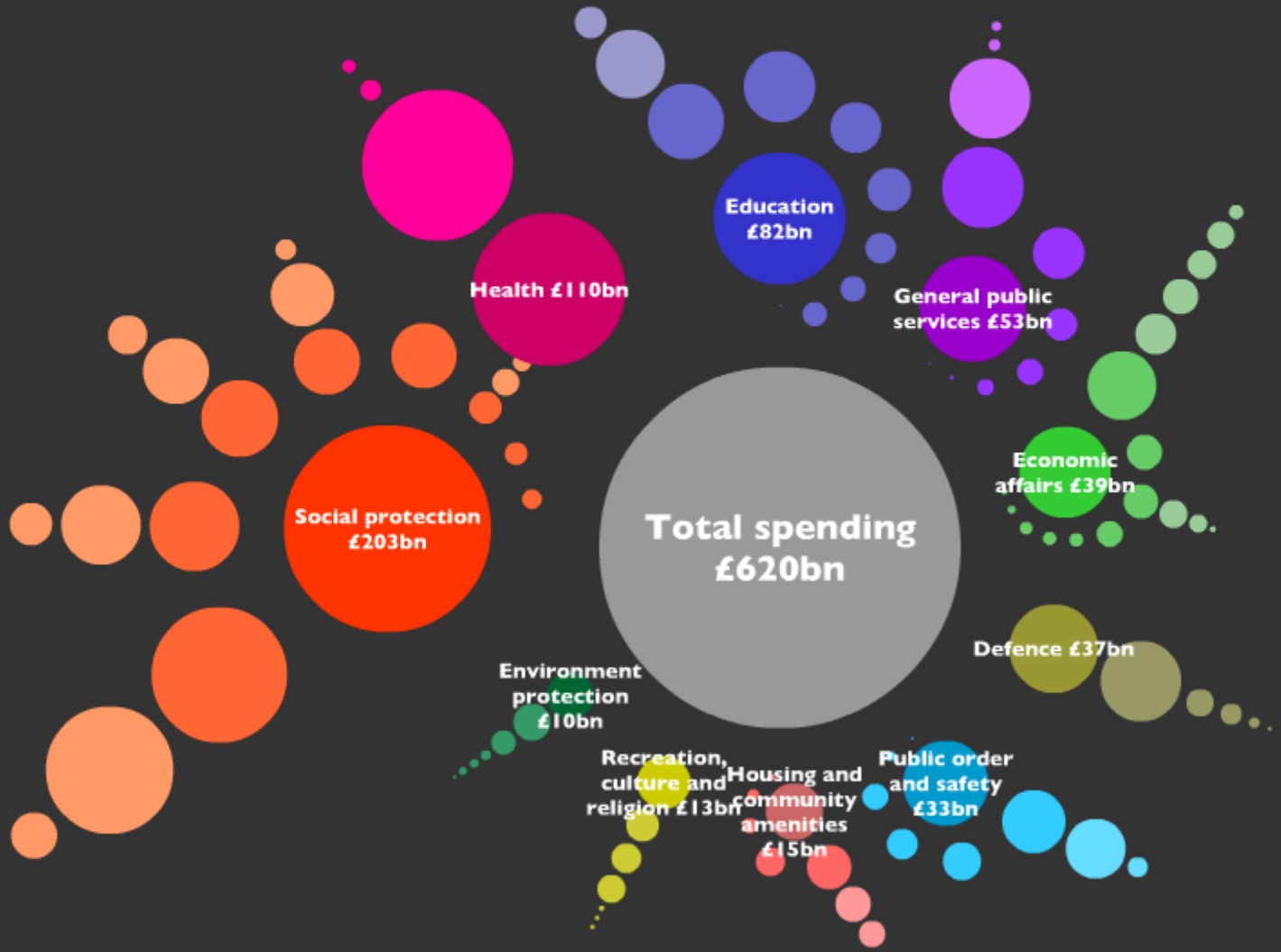
Where Does My Money Go?

UK WIDE | NATIONAL | REGIONAL

Overview | Long-term trends ▶

- Economic affairs
- General public services
- Education
- Health
- Social protection

2003-04



For more information about my work please see the Open Knowledge Foundation website at:

http://www.okfn.org/

Image credits

Work with schools by New York Public Library
Participants by jwyg
Plumbing Bits by cmurtaugh
Harwell Dekatron Computer by MichaelWilson78
Pierre Vivant's Traffic Light Tree by William Warby
The Green Light by Ted Percival
Online does not equal open by jwyg
Lego Bricks by bdesham

These slides are available under a Creative Commons Attribution Sharealike License.