Bootstrapping the open data ecosystem

Dr. Rufus Pollock
Fellow in Economics
University of Cambridge

Director
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
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 many more!
Confusing Signals
The best thing to do with your data will be thought of by someone else!
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

Unlocking innovation
Working with UK Public Sector information and data

Advised by Sir Tim Berners-Lee and Professor Nigel Shadbolt and others, government are opening up data for reuse. This site seeks to give a way into the wealth of government data and is under constant development. We want to work with you to make it better.

We're very aware that there are more people like you outside of government who have the skills and abilities to make wonderful things out of public data. These are our first steps in building a collaborative relationship with you.
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.

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
format-csv  geodata  format-xsl  access-api  license-not-specified  music
ckanupload.esw.200910  format-pdf
gov  size-large  format-xml  environment  genetics  development
access-bulk  aidinfo  country-uk  publicdomain
glen  science  bibliographic  city  access-search  workshop-20081101  language
uk  chemical  lod  country-usa  images

format-rdf  government  climate  economic

Recently changed

<table>
<thead>
<tr>
<th>Packages</th>
<th>Author</th>
<th>Time</th>
</tr>
</thead>
</table>
The "Debian" of Data
Automating use and reuse of data

The CKAN System

CKAN (The Registry)
+ Web User Interface (WUI)
+ Machine Web API (WAPI)

Related Services

Is It Open
isitopendata.org

API (JSON)

CKAN CLIENT

STORAGE (The Payload)
archive.org / knowledgeforge.net repositories (svn/hg/git/...)

DATAPKG
(like apt-get,gem,pip)
Where Does My Money Go?

Total spending £620bn

- Social protection £203bn
- Health £110bn
- General public services £53bn
- Education £82bn
- Economic affairs £39bn
- Defence £37bn

Where Does My Money Go is an Open Knowledge Foundation project. Source data...
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.