OECD Foresight Forum on Big Data

22nd Oct 2012

Rufus Pollock

@okfn{.org}
@rufuspollock{.org}
Open Knowledge Foundation

@okfn[.org]

We build technology and communities to develop, disseminate and use open information — material that everyone can use, share and build on.
Open vs Closed
Anyone means anyone - no restrictions on commercial use. (So e.g. CC NC licenses are not open!)
Most (Big) Data is Closed

- Value in keeping proprietary
- Privacy issues (in theory ...)
"Info about people or things associated to people"
Opening up Big Data
Making the most of its potential

"The best thing to do with your data will be thought of by someone else"
Is small more interesting than big?

What an individual can do today on their own (perhaps on their laptop) - not what IBM/Oracle/... can do for you

Small pieces, loosely joined

vs

One Ring to Rule them All
Small vs Big
- Is small more interesting than big?
- What an individual can do today on their own (on a laptop) - **not** what IBM/Oracle/... can do for you
- Small pieces, loosely joined vs One Ring to Rule them All

UK Nat Gov Spending for last 2y = ~4.1m items

All of it on my laptop today

Monoliths versus networks
Most Examples Today (Info on people or things assoc to people)

Anonymization, Sampling, etc
Build them (& open publication) into the process (+ retain IP with gov / users!)

Currently few examples (gov)

MyData

Sao Paulo metro
Each day: 6m people, ~24m journeys => 720m per month

Random sampling and anonymization => 1-10m records "small" dataset which can be open

Rapidly growing (gov etc)
Conclusion

Many issues of big data are general issues of data (e.g. privacy)

Big data not a stable category (today's big is tomorrow's small). Small may be more interesting than big

Other categorizations may be more useful - personal vs non-personal, closed vs open, real-time vs historical

Ensure access & (wherever possible) open access for gov, for citizens

Open vs Closed is a "bigger" issue