

Mapping Corporate Networks and Public/Private Spending Using Open Data

Tony Hirst, @psychemedia, tony.hirst@open.ac.uk

Affiliation: 1) Dept of Communication and Systems, The Open University. 2) OKF

User context: open corporate data and: 1) investigative data journalism; 2) academic research; 3) transparency.

Technology context: data representation; data linkage & reconciliation; data modeling for storytelling.

For some time, I have been exploring ways of using open company data from OpenCorporates as a focus for a range of technology experiments and learnings:

- exploring 'natural' representations for datasets, particular the use of graph (network) representations; eg mapping 'corporate sprawls' based on co-director networks, and registered address similarity;
- representing data in 'human meaningful' form; eg generating galleries of widely recognised trademarked logos that are associated with particular companies (consumer empowerment – “I didn't know company X made products A and B?!”)
- entity reconciliation; eg reconciling companies listed in open spending data with OpenCorporates identifiers, with a view to assigning them to corporate sprawls;
- cross-domain mapping; eg looking up directorships of local councillors and comparing them to companies those councils spend with; looking up company details associated with companies making planning applications in particular areas.

Whilst working with this data, several issues have emerged, (including technical issues that are likely to be encountered in other domains):

- lack of unique identifiers for entities, eg directors (is 'this' John Smith the same person as 'that' John Smith);
- variety in addresses, eg registered addresses for companies (should we be geocoding addresses as the basis for address searching/discovery?)
- varied and un-normalised access to information depending on geographical locale, eg in OpenCorproates. sparse access to director information in many territories outside the UK;
- handling the reporting of groups and relations between group entities; eg no clear way of reporting, aggregating and identifying separate companies that are part of the same 'group', nor the linkages between them (should we be using graph based representations?)
- lack of APIs and open data for ascertaining ownership; eg around trademarks (eg machine searching for trademarks owned by a particular company).

One workflow I currently use is based on using Scaperwiki for running structured searches over the OpenCorporates API, compiling my own thematic data views therein, expressing the data as a graph/network using 'standard' JSON and XML representations, and then rendering those network graphs using standards accepting Javascript libraries and desktop tools.

I am happy to report/present on my work in this area to date.

At the workshop, I would be keen to explore issues that could help further my thinking around the following issues:

- 1) ways of generating and reporting linkage within OpenCorporates data or other open corporate data datasets;
- 2) ways of generating linkage in to and out from OpenCorporates - for example, reconciliation services based around: i) company names; ii) director names; iii) corporate registered addresses. Such linkage may come from Companies House, openlylocal (companys, councillors), opencharities (trustees, addresses), openspending (companies), MP declaration of interests (companies), All-Parliamentary Group sponsors (companies) etc (the latter relating to the construction and identification of lobbying networks).
- 3) ways of representing linkage between groups of elements in corporate and person/corporate networks.

I would also be interested in discussing methods generally that might inform ways displaying the evolution of corporate and director networks over time, along with timelines based on company formation and dissolution data, as well as director appointment and termination dates. (For example, one use case might be detecting companies that are quickly set up and ripped down; another might be looking for companies that are dissolved soon after the appointment or termination of a particular director).

Examples:

A "corporate sprawl" based on co-directorships:



Local council spend with elements of the same corporate sprawl:

