Assembling Open Data Complementarities for Service Innovation

Abstract

Strong claims are made about the potential of opening government data to drive service innovation. Yet little is known about the detailed processes of how hackers create services out of releases of public datasets, and the conditions for the move from data release to service innovation. In a multimethod study of the open data hackers in the UK we identified a series of interlocking processes involved in the conversion of public data into services of public value. We found that few of the 'rapid prototypes' developed through hack day events are maintained or sustained as service innovations beyond the hack days. We identified key artefacts involved in the realisation of complementarities: cleaned data available through APIs or bulk downloads, linkable data, shared source code and configuration, source code repositories, existing digital services. Our inductive findings were checked through a statistical test on the predictors of complementarities, confirming a range of hypotheses on the motivations and behaviours of hackers that support sustainable use of open data for service innovation.

Keywords: *Open data, complementarities, sociomateriality, motivations*

The release online of government datasets, under open licenses allowing for re-use, has received sustained policy attention under both previous and current UK government administrations. Since the data.gov.uk data store was publicly launched in January 2010 with nearly 3,000 datasets listings, the range of UK open government data available for re-use has grown significantly, with pressure on government departments, local government and trading funds to release raw data. Arguments given for the release of data include reference to increased transparency and accountability, and the potential for open government data to stimulate the development of the web of linked data and the semantic web. One of the most prominent arguments for the release of data, however, is that developers and hackers will be able to take raw datasets and create mash-ups and applications, combining public data with other datasets and services to provide useful information, tools and services to the public. This is seen as supporting both co-production of public services, and the development of service innovations, as entrepreneurial actors outside the state bypass complex procurement arrangements and institutional barriers to deliver prototypes and scalable services based on open data.

High profile competitions such as Apps for Democracy in the US, and Show Us a Better Way in the UK have emphasised the potential benefits (both in terms of financial savings to the state, and potential for innovation) of releasing data to developers, and allowing actors outside of government to build products and services off the back of it. Both

Data.gov.uk and the London Data Store feature ‘Apps’ or ‘Inspirational Uses’ of the data they provide - ranging from new websites that provide interactive access to performance data on schools, to visualisations of travel times in London, and a Live Tube Map, drawing on London Underground data to show the location of underground trains on the network. Whilst these anecdotal examples underline the intuitive plausibility of the claim that opening access to government data can bring significant benefits through third-party hackers working with it, established and systematic metrics for evaluating the impact of open data are lacking.

Tracking the impact of open data is complicated by the difficulty of tracing the value-chain for data which, by its nature, required no formal relationship between supplier and user (Dekker et. al. 2006). Rather than focusing upon establishing outcome measures (e.g. counting the number of new start-ups building around data-driven businesses), this paper examines key intermediaries that complement open data and notably enable the process of change (Robey and Boudreau 1999). By tracing the different processes and sequences involved in the use of open data for provision of information and services of public value, we identify elements of those processes which promote or undermine the realisation of that value. Our central argument is that the utility of open data can be enhanced through complementarities. Data (even timely and comprehensive data) is, on it’s own, insufficient to drive innovation in the design and development of better services. Our findings highlight the different human and material

agencies involved in effective use of data, and show how effective processes of open data use draw upon, and contribute to the development of, open data complementarities.

To develop our theory, we used a combination of methods comprising ethnographic research, archival analyses of blogs and activities of hackers, interviews with key informants and statistical analysis of survey data. While we collated and analyzed our data, we identified theoretical categories that led us back to the literature. This approach provided a practical middle ground to iterate between empirical data and theory, avoiding the pitfalls of ignoring the literature (Suddaby 2006). During our analysis, we identified the theory of sociomateriality (Orlikowski and Scott 2008) as an appropriate lens. For clarity of presentation, we review the literature prior to the discussion of the research setting and our findings although we examined it throughout this research as data collection and analysis indicated its theoretical relevance. Following our discussion of the literature, we provide more detail on the methods, and then present findings in two sections: the first presents qualitative inductive findings from the data, and the second presents a deductive statistical analysis to confirm and extend the inductive findings, focussing specifically on the connections between motivations for sustaining open data engagement and the accumulation of complementarities.

Complementarities and agency

The notion of complementarities has been variously defined and used in different contexts. A dominant theory in the literature is the economic theory of complementarities. The theory is

often understood in terms of super- and sub-modularity (Tanriverdi and Venkatraman 2005), in that, coordinated rather than uncoordinated activities yield not only higher returns but also lower cost. It suggests that certain activities when brought together are more than the sum of their parts, in that, coordinated activities are mutually enabling and reinforcing each other's performance. The theory is implicitly limited to being a contingency account, as if coordinated activities provide the only optimal condition for realizing complementarities (Cooper and Haltiwanger 1996). Furthermore, the implicit assumption that both coordinated activities can exist independent of each other has precluded a closer examination of the emergence of complementarities.

Analysis of our data suggested there are other conditions (including contextual, e.g. Why coordination happens in certain tasks (or datasets) but not others?; intervening, e.g. Who and/or what instigates the coordination?; and causal, e.g. Can one activity lead to the emergence of the other to form one complementary set?) that explain why a contingency approach towards open-data use and service innovation can fail without accounting for two key elements: temporal emergence and sustainability of complementarities.

Andrew Pickering (1995) used the concept of mangle to underscore the significance of temporal emergence in practice, suggesting that the way human and technology are co-configured through social actions can either dissipate or generate resources in an on-going manner, and that the resulting intermediaries (including new roles, routines, and material

solutions) temporarily emerge to overcome the material constraints. His concept has been sanctioned by other IS scholars (Jones 1998; Orlikowski 2006) to account for how the materiality of technology emerges, are sustained and configured. This practice-based perspective provides a useful lens to study the conditions (contextual, intervening, and causal) that can induce the assemblage of complementarities.

Any analysis of technology risks falling either into over-prioritising the role of technologies (material agency), or over-prioritising the role of human agents and agency in realising a particular set of outcomes. A sociomaterial lens draws attention to the mutually constituted nature of both human and material agency (Orlikowski and Scott 2008), and the roles that social and material artefact play. Volkoff and her associates (2007) use a longitudinal case study to show how roles and routines were inscribed in technology through instantiating the material aspects of technology, which in turn circumscribe other actors' future activities. Mutch (2010) argues that, "we need to consider not only the ways in which structures can be embedded in technology and at which levels, but also how such embedding is perceived by and responded to by a range of users" (p. 511). He also argues that "in many circumstances, structures, language being a key one, are bequeathed to us by actors no longer present, but they form the involuntary context, a context that can both constrain and enable, for our action now" (p. 510). In relation to open data, the ways that data has been collected and the formats it is encoded in may impose a set of constrains for further use, as they are collected for different

administrative purposes (such as census, accounting, etc); and that, for example, they may appeal more to the press and right activists than the hackers.

In the case of open source, social norms and culture provide key context, but material infrastructures and artefacts support its instantiation in action. Many hackers are incentivized by (in addition to private and social reasons such as fun, peer recognition, and showcase of abilities; see von Hippel and von Krogh, 2006) ideological commitment to the idea of freedom to reuse and modify codes. The extent to which codes are malleable for individual and collective use provides a unique set of opportunities for individual learning and collective creativity. Lanzara and Morner argue that "the source code and its multiple beta versions anchor and circulate software-related knowledge while at the same time inscribing relevant components of human agency and social interaction that facilitate the coordination of a high number of human agents" (Lanzara and Morner 2003, p.3).

The source codes serve to provide a structure to cement and/or facilitate actions, and interactions of hackers. Hackers can reuse each other's codes in the public repository by combining with new ones in developing further modules and files of the same and/or a different app. The app represents the material agency and the actions of modifying codes and developing new apps represent the human agency. In contrast to human agency, Leonardi (2011) defines material agency "as the capacity for nonhuman entities to act on their own, apart from human intervention. As nonhuman entities, technologies exercise agency through their

performativity", through which the utility of material agency is realized. He illustrates when human and material agencies are interlocked in a particular sequence, the interaction can be mutually enabling and initiate a set of conditions that are conducive to change.

To further explain how complementarities can be developed despite the constraints that often come with open data, we collected both qualitative and quantitative data to examine how open data were being assembled and reassembled by hackers. We identified several interlocking sequences of this process centring around the creation, or use, of five distinct artefacts: cleaned data, linkable data[1], software source code, **s**hared source code (in a revision control system such as git hub), and service technologies. We found that the 'hacky' nature of much current open government data (including missing data, poorly curated data, tricky to use formats) spurred concreted effort of cleaning by hackers. The cleaning increases the reuse value of open data by providing a linkage point to mash up with other types of data. This not only reinforces the value chain of data to information but also induces the creation of new software and/or the use of analytics and visualization tools. By making source code accessible through a source code repository this induces further code modification in the support of developing new service technologies, which ultimately supports the development of new services, and/or the integration of data into existing services.

[1] N.B. This is not necessarily 'linked data' of the form described by Burners-Lee (2006), but it at least data which is easy to connect to other datasets by use of standardised identifiers and shared keys. For example, using established Office of National Statistic (ONS) codes for identifying local areas.

This research provides two distinct theoretical contributions. First, we show how artefacts provide a linchpin for human and material agencies to work in tandem. This dovetails with prior research, in that, by separating artefacts from material agency, the role of artefacts serves to complete human agency, and at the same time, either permits or precludes material agency. Although both artefacts and material agency collectively define the materiality of technology, the decoupling explains why complementarities temporarily emerge or fail to emerge (e.g. complementarity is restricted when source code, or cleaned data, is not released to the public and instead remains proprietary). Second, we clarify the temporal emergence of complementarities is a sequentially interdependent process, with which human and material agency recursively constituting each other, mediated through the creation of artefacts. The five artefacts we identify create a stack of complements similar to that of software stack (Gao and Iyer 2006), each stack layer either enables or constrains further development. This observation suggests that ways in which complementarities of open data are assembled and/or reassembled are more recursively interdependent than linearly independent.

Methods

In attempting to tackle an emerging phenomena, we used multiple methods as a way "to attack a research problem with an arsenal of methods that have non-overlapping weakness in addition to their complementary strength" (Brewer and Hunter 2006, p.4). To provide an overview of the development in open data field, from January 2010 to June 2010 we used a

custom-built computer assisted qualitative data analysis system to record and analyze public Twitter messages including the '#opendata' hashtag (Huang et al. 2010), presenting data in tag-clouds for exploration (Rivadeneira et al. 2007). Tweets were regularly reviewed and emerging themes recorded through a private wiki-based research journal (Borg 2001).

Between March and October 2010, we participated in 7 open-data events which involved discussions of policy and the challenges in using open data including datasets made available by Ordnance Survey, UK Government Data Store, London Data Store and Open Street Map. These events also included hands-on hack-days, working with public data to generate innovative prototypes and services. Table 1 summarizes the events and their respective aims. Participants involved policy makers, data managers, commercial organizations, and hackers/developers. The events provided us with a fresh empirical perspective not bound to the existing theories, deepening our understanding of issues related to open data use and facilitating the identification of key questions and themes for later analysis.

INSERT TABLE 1 ABOUT HERE

We also draw upon cases of open data use described in the survey component detailed in the deductive method below. The survey data, ethnographic observation and interaction with open data users led us to focus upon the documentary evidence and participants from the first ten hack-days (between March 2009 and December 2010) organized by Rewired State (http://rewiredstate.org/), allowing us to gain a deeper understanding of the practical processes

by which open data was used. Rewired state described these events as "where developers show government what is possible, and government shows developers what is needed". The Rewired State website records details of 'hacks' created at the one or two-day events it organised, and allowed us to identify the datasets attracted most attention, and the identities of the lead hackers. In the ten hack-days, a total of 130 open-data projects were initiated, but only 43 document producing clear prototypes. Of these only 10 remained active at the start of 2011 (based on an identification of those remaining updated and maintained).

To supplement the field and archival data, eight semi-structured interviews were conducted. Interviews invited open data users including hackers to give detailed account of a particular instance of open data use, their reasons for working with open data, and challenges they encountered. All but one interview took place by phone, recorded, transcribed and coded for key themes. By triangulating among multiple sources of evidence, the multimethod provided greater depth and accuracy.

INSERT TABLE 2 ABOUT HERE

With the profiles, blogs and tweets of project leaders and contributors of hackdays, and a range of uses of data described in survey results, our analysis proceeded through four rounds of coding. In the first round, we coded data use instances into five types as summarized and illustrated in Table 2. In the second round of coding, each author independently identified the reasoning and rationales for the particular uses of open data, and expressed challenges and

issues relating to the use of open data. We then replicated the coding with transcripts of the interviews. In the third round of coding, we wanted to determine how open data were assembled and reassembled to increase use and reuse. Thus, we studied the ways hackers appropriated and used open data, and the types of artefacts and technical solutions that they produced. We identified five archetypes of artefacts: cleaned data, linkable data, source code, shared source code repository, and service technologies; and five aspects related to technical solutions: application programming interface and/or cleaned data dump; mashups and applications; source code and configuration; shared source code repository; code library and configuration; and service integration. In the fourth round of coding, we sought to discover relationships in the data, by coding for conditions (intervening, causal, and contextual), and actions and interactions, and consequences. We replicated the coding with two further sets of embedded cases of open data use from the London Data Store. The replication of the coding on embedded cases was to subject our analysis to further testing (Yin 2003), ensuring that emerging themes and findings were not specific to a specific dataset, event or medium of interaction. The first set focused on education data (the EduBase dataset); the second set on public spending data (the COINS).

Inductive Results

The presentation of our results in this section is organized around three core elements: the imperfect 'hacky' nature of much current open government data; the motivations of hackers

and other open data users; and the assemblage of open data complementarities through complex process of open data use. We argue that the emergence of complementarities is principally rooted in the motivations and culture of hackers as users of open data, responding to the constraints on the data available to them. Open data is seldom sufficient on its own to induce hackers' involvement, and the ways hackers modified and reassembled open data introduce a material aspect to the resulting technology which either constrains or enables further development.

Open Data

Although considerable quantities of data have been released through portals such as data.gov.uk, the quality of that data varies widely. Hackers we spoke with unanimously noted the poor quality of much open government data, ranging from bad data formats (e.g. "*issues with quotation marks in the CSV data which can be confused with field delimiters, caused errors in mapping the data to XML*") and infrequent releases, to a lack of granularity or inconsistency in naming or choice of identifiers. Hackers contested some open datasets are "*no more than a bunch of statistics not useful whatsoever to derive any utility*" and some datasets, often Excel datasets which rely heavily on visual presentation, are "very pretty looking things" but that have little re-use value. Another hacker also commented that:

> Applications based on old data which is out of date are worthless, they don't provide utility at all, and they only build customer dissatisfaction.

Yet the imperfect nature of open data did not deter many hackers from working with it. A lead hacker disclosed that:

> *The secret to our success was not staring at the endless list of incomplete/hacky data but actually asking ourselves what the government was doing wrong*.

The incomplete/hacky nature of open data provided a degree of malleability for hackers to clean the data, by "*chucking out all the dodgy stuff because there is some dodgy stuff in there*", in order to provide useful and usable data.

Hackers

We explored uses of open data in a range of contexts, from hack-day events, where hackers come together in one location and set themselves the challenge of initiating a new open data-related project in one or two days of intense activity, to uses of open data by individuals working in their own time to explore and engage with data. Because our period of study has been one in which a lot of new datasets were released as open data, even without hack-days being organised, the periods shortly following to release of key datasets such as the COINS public spending dataset, and the Ordnance Survey postcode databases, often took the form of 'virtual hack-days', with hackers across the country working on the same datasets in loosely co-ordinated networks.

There are many different factors that incentivize engagement with open data. Both hackers and non-hackers can engage with data. For our purposes we define hackers as individuals capable of making advanced use of digital tools, or writing code, in order to work

with, manipulate and generate new artefacts with datasets. Analysing a wide range of instances of open data use, and drawing on conversations with and documentary evidence from open data users, we identify a range of factors that incentivize engagement with open data.

For some, desire for access to specific facts motivates engagement with data. One open data user noted that as he had lost one appeal regarding a school place for his daughter, so he decided to look up for the appeal data:

> I was just interested in trying to find out how many appeals succeeded and really what the status was around the different councils and whether we actually had a chance...when you are in that sort of situation you want to know about what happens to other people and how likely it's, and the type of arguments that do win.

In this instance, once the user had located the facts he was looking for, he had no incentive to develop tools or applications to help others access this data, nor to conduct any further analysis of it. By contrast, a school governor challenged by the education authority over his school's revenue balance (retained funds), sought out multiple sources of public data (including data obtained through a Freedom of Information request as it was not proactively published) and combined them in an analysis published on his blog. The details of how the analysis was carried out were not published (though they are detailed in E5 based on an interview with the individual), but rather the information generated was offered as an advocacy tool for debate over revenue balance policy. The governor noted that:

> "These numbers tell a story, and if you go out and say I want to reduce all the other balances you may end up finding that you're actually dis-improving the schools that are out there because generally it's a characteristic of good schools to have a high balance."

Some hacker users of open data were driven by a specific desire to show how government services could be run better or more efficiently with digital technologies. The tag-line of Rewired State ("Coding a better country") captures this idea. A number of instances of open data use were specifically oriented at demonstrating what could happen if *more* data was opened up – highlighting the potential of data, and the current limitations on it's realisation given the limited support of data.

An elder hacker noted that "*younger hackers were more interested in solving problems*"; and that once they identified the problems that needed solving, the next logical step for them is to consider where to "*source the right data*". This may contrast with those with greater awareness of current data availability, and so who focus their energy on problems tractable with currently available data, including recently released data.

As with the fact- and information-focussed uses of data above, it was common for hackers to draw on personal frustrations as a motivation for development. Often this might include trying to identify a shared problem and then to focus on sourcing data or cleaning hacky data in order to deal with it. One younger hacker noted that:

> Everyone, at some point in time, has been irritated with the excessive demand on select tube stations – anyone that's been past Oxford Circus in rush hour will understand where I am coming from. In addition, time is at a premium in London more than anywhere else – every minute lost waiting around for a train (or worse, waiting to get off of one) is a minute's money. Our solution is to show people hotspots, and how to avoid them.

In other data-use instances, we observed the utility value had direct relevance and appeal to hackers' current situation and beliefs, e.g. "*moving to London* [hence contributing to this particular project]", "[this app] *helps you to get fit*", "*I am a bike user*...[this app] *can get people to move bikes to emptier docks*". In some situations, hackers developed solutions to directly benefit their collaboration with other hackers as illustrated by the following quote:

> We stumbled upon the fact that we'd been using the same set of tools for the past few hack days, and then the idea hit us that we could build a tool which just helped to coordinate rapid development…after all, regardless of whether the project won or not, we would still end up with a great tool we could use in the future.

We observed that open data users were motivated for not only by altruistic, private and social reasons but also by the prospects of monetary reward, and engaging with the economic potential of open data. One hacker noted that:

> M y two hats are very different. Certainly, as a private citizen I see the way people have adapted and started using the data...on a professional level, we will see our data actually being published in things like RDF in the not too distant future to allow it to be found and used more easily, and then we have to figure out how we make money out of it in time, and keep me in employment.

Many of the hackers involved in hack-days developed software as a profession rather than a hobby, though not necessarily working with public sector data on a day-to-day basis. For some hackers, their contribution to open hack-day events aimed to showcase their abilities to prospective employers. A hacker unreservedly stated "*offer me a job*" on his website. And among all the lead project developers of the Rewired State projects we explored, 78% provided

their affiliated work contacts (including names of the companies and/or email addresses), with others consisting of young hackers still in education, academics and researchers.

Although over 3000 datasets had been released from data.gov.uk, and more from local data stores such as the London Data Store, we observed hackers clustered around specific datasets and APIs, partly due to the completeness of the available data, and partly due to the thematic focus of that data. This pattern was also particularly pronounced exploring hack-days related to the London Data Store. Although the London data store included over 425 datasets at the time of study, we found just 47 uses of that data publically noted. Amongst the datasets, 8 in particular attracted the majority of attention – all of which were transport-related datasets. We noted 22 instances of this data being put to use, generating 8 smart phone apps (4 paid, 4 free), 2 APIs (application programming interfaces) and 12 websites (including interactive maps, graphs and visualisations). Whilst some were 'experimental' uses of the data, a number of these would constitute clear service innovations – creating tools that the state had failed to provide with prior proprietary access to the data. The disproportionate focus on particular datasets can be attributed to the fact that hackers developed solutions in response to the popular demands from the public. On the website of London Data Store, the public was invited to make suggestions and vote for their popularities. The top ten mostly voted suggestions were all transport related. Hence, it was not surprising that the eight transport datasets attracted most developers' attention, and that 78% of the solutions were contributed by hackers (either

working for themselves or commercial organizations) and the rest were contributed by researchers affiliated to academic/research institutes. The utility accounts for hackers' preferential attachment, in that, hackers were most likely to provide solutions that yielded the most utility to themselves and/or the public. This has interesting implications for identifying the datasets around which complementarities may autonomously emerge, and those datasets (more niche with respect to hacker interests, but potentially of significant social value) around which active effort may be required to assemble complementarities that support their re-use in service innovation.

Assemblage of complementarities

Direct use of open data seldom occurred in most of the data-use instances that we analyzed. We found open data had to be first cleaned or curated by hackers, and then made available as a cleaned data dump on a website or through the use of API. The cleaning process changed the material aspect of the original dataset, and created a new artefact that permitted the use of a technical platform for its distribution. The technical platform effectively performed the role of material agency by distributing the cleaner version of open data. The cleaned data which was structured and discoverable permitted various linking possibilities with other datasets (e.g. linking against schools/departments, programme codes, places, and people). The resulting linkable datasets offered other open data users the opportunities for exploiting some of the known analytics and visualization techniques. Others involved hackers writing code

and/or configuring existing tools to work with the linkable datasets, for example, configuring Yahoo Pipes by drag and drop.

By separating artefacts from material agency, this allowed us to identify the key intermediaries in the assemblage of open data complementarities for service innovation. The artefacts formed the components of a value stack: from messy raw data to linkable data; from linkable data to software; from software to shared code repository; and from repository to new web services. At each layer of the stack, the material aspects of the artefact served to linchpin human and material agencies in an interlocking sequence. We identified five interlocking sequences to illustrate the emergence of complementarities.

INSERT FIGURE 1 ABOUT HERE

In the first interlocking sequence, human agency changed the material aspect of open data (by cleaning the data) and created a cleaner dataset (artefact), and the material agency enhanced the performativity for easy access and distribution of the data through the use of API (or data dump of the cleaned data). In the second interlocking sequence, the material aspect of the cleaned dataset (artefact) intended to improve the understanding of the data, in that, not only what users could do with the data but also how they could draw in extra data, without need for prior knowledge of all other APIs or data sources. With various possibilities of how cleaned data could be linked with other usable and useful datasets induced further exploration activities (human agency) to improve the contextual understanding of the data. This was

similar to the two aforementioned data-use instances regarding the rate of successful appeals, and the relationship between school performance and annual reserve.

In the third interlocking sequence, despite its intended goals (material agency), linkable data could still impose certain constraints and challenge users to look for social and technical support. The following two quotes illustrated the limits of linkable data, and how human agency sought to circumvent the material constraints:

> I put a plea out of Twitter saying anyone knows where I can get political controls of councils and someone pointed me to a page somewhere. It's some with a geeky interest in it who has actually gone out and got the data.
> We reviewed as many of the provided datasets as possible, looking for common data points such as people, events or places on which we could do a linking exercise. We drew a blank on this, and instead did a mapping exercise based on NGR co-ordinates.

In addition to seeking support, other open-data users would reconfigure existing tools, and write codes for data mashups and new applications. This marked the beginning of the fourth locking sequence which involved the hackers. Hackers who embraced open source would chose to deposit the source code of their apps in a shared source code repository. The donation underlined the ethos of open approach that "*it won't be a sole developer who ultimately brings it to fruition*". The private donation by hackers led to the last interlocking sequence whereby other hackers could reuse and modify the codes, in a limited number of cases we observed, integrating shared code and/or open data into new service technologies, or using it to complement and innovate within an existing service.

INSERT TABLES 3 AND 4 ABOUT HERE

Because the data use instances were drawn from multiple datasets, we applied the framework of interlocking sequences in two sets of embedded cases to show the sequential interdependency of the value stack. Tables 3 and 4 illustrate how same dataset can lead to different assemblage of complementarities, and how, in a number of cases, the absence of certain sequences and complementarities has led to duplication of efforts. For the embedded cases of education data-use, the cases E1, E3, E4, and E8 exhibited the characteristics of interlocking sequences 1, 2, and 3; E2 exhibited 1 and 2; and E5, E6 and E7 exhibited the first interlocking sequence. And for the embedded cases of COINS data-used, only C4 exhibited all five interlocking sequences; C2, C5 and C6 exhibited 1, 2 and 3, and C3 exhibited 1 and 2.

The case C4 describes the Open Knowledge Foundation (OKF) use of COINS public spending data, and exhibits all five interlocking sequences. It is worth noting that there was close collaboration between the hackers involved in C3 and C4, although not represented in the diagram. On the day of COINS launch, OKF collaborated with others to clean the data, using an etherpad and IRC channel to discuss strategies for cleaning the data. The data was imported into their own database, and they provided an API onto it, allowing others to draw upon the clean data in raw or filtered form. As the COINS data made use of a lot of programme codes and categories for which no public code-list was available, they worked together with The Guardian Newspaper to crowd-source definitions of codes, improving the 'linkability' of the data. The processed and linkable spending data has subsequently been used to generate a range

of visualisations and analysis, and to power WhereDoesMyMoneyGo.com, and existing open source project for visualising spending data that has been adopted as a framework by a number of other national projects working to visualise public spending. Source code for working with COINs, and for the WhereDoesMyMoneyGo site has been shared through the OKFs own source code repository.

Conclusion

To understand what complements open data, we examine the underlying processes that constitute the conditions for physical and social complementarities to emerge and accumulate. Our findings suggest the initial efforts by hackers (driven by the desire to bring utility to themselves and/or the public) set the momentum for change. The cleaned data provides the first key intermediary for material agency to follow through the use of an open API or data dump. This induces further actions to enhance the utility of open data through the creation of an additional set of complementarities. Embracing open approach, hackers share the codes of their apps to induce further chain of actions and interactions. The interlocking sequences between human and material agencies, and their interactions with artefacts create the conditions for change; and notably accelerate the development of innovative solutions.

Our theory suggests complementarities of open data are sociomaterial assemblages (Latour 2005; Markus and Siliver 2008), constituted and re-constituted through an on-going interchange between human and material agencies, and during which new artefacts are created

and new technical solutions developed. Our theory can be used to derive evaluative frameworks to better assess the impact of open data at different stages of it's use – both in terms of measures of tangible (such as new datasets) and intangible assets (reputation effect of hacking) along the path from the releases of open data to the creation of innovative services.

Part 2: Deductive Approach

Our qualitative findings have revealed different motivations for contributing to open data projects, with initial (or short-term) participation often characterized by use-value motivation and with extended participation driven by the anticipated rewards including developing innovative web services and technologies, and/or monetary gains. From our qualitative data, it would appear that anticipated rewards contributed more to the stack of complementarities than use-value motivation. To test this prediction, we used survey data to statistically model the impacts of different types of motivation in the use of open data on complementarities.

The data-use instances demonstrated the innovation possibilities, but rarely were they sustained or developed into sustainable services. Of the 43 initiated projects in the ten Rewired State events, only ten remained actively updated and maintained. The active projects exhibited several unique characteristics which were similar to a typical open source project (e.g. Kuk 2006). They comprised: not a loner project; having immediate relevance and appeal to the hackers; devising a technical solution to a well-defined problem; aiming to form an open

source community; seeking to improve the reuse value of data and other associated artefacts; and seeking to exploit the resulting technologies for service innovation and/or profit. Whereas other non-active projects were characterized by short-term goals, i.e. using open data to solve a problem of personal needs and use benefit (use value).

In the context of participation in open source projects, Shah (2006) found in contrast to long-term participants, short-term participants were driven by use value. Bagozzi and Dholakia (2006, p. 1111) also found that contributions by inexperienced hackers are "driven by specific task-oriented goals typically of a personal nature". This was exemplified in one of the aforementioned comments made by a young hacker. Roberts et al. (2006) found that different motivations affected participation in different ways, in that use-value driven participants contributed significantly less than paid participants.

As software is an integral part of open-data use, we expect motivations matter in terms of the level of resources (including time, knowledge, and skills) that hackers will commit to maintain the open-date projects. That is, hackers who anticipate future rewards of open data in terms of service innovation and profit are more likely to sustain participation, which in turn contributed to build the stack of complementarities. Whereas hackers who are use-value driven are likely to use open data as a means to solve a specific problem in their short-term participation. The above discussion led to the following three hypotheses.

HYPOTHESIS 1. Hackers who are motivated by the anticipated rewards of open data are likely to sustain their participation

HYPOTHESIS 2. Sustained participation is positively related to building complementarities.

HYPOTHESIS 3. Hackers who are use-value driven are likely to positively perceive data constraint as a challenge than a threat.

Hackers who are extrinsically motivated are also likely to perceive their participation in open-data projects having a positive impact on society and communities. Roberts et al. (2006) argued that extrinsic motivation can work in tandem rather than crowd out intrinsic motivation. This follows the trend that most of the hackers often write code as a profession, and at the same share code with others as a hobby (Fitzgerald 2006). This explains why hackers can work under mixed incentives because the sharing of codes goes towards enhancing the hackers' status within a meritocratic system as in open source communities (e.g. Lee & Cole 2003); and employment prospect (Lerner & Tirole 2002).

HYPOTHESIS 4. Hackers who are motivated by the anticipated rewards of open data will perceive their contributions exerting positive impact on society and communities

On the public launch of data.gov.uk, a tweet by an activist noted that "open data is not a panacea, but it's a start", advocating for more efficient and honest government. From our most embedded cases, although we could not identify their direct social impact, they were connected with public sector reform and/or democratic change. With the use of COINS, we found that conventional forms of scrutiny (C4, newspaper led) complemented by crowd-sourcing initiatives (C3-C4) and independent individuals (C1, C5) and companies (C2) providing their own platforms for members of the public to explore government spending, and to scrutinize that spending. The availability of multiple education datasets also allowed an independent

school governor (E5) to scrutinize a specific government policy. This led to the following hypothesis.

> *HYPOTHESIS 5. Hackers who are driven to make government more efficient and accountable are likely to perceive that their involvement in projects will make a positive impact on social welfare*

Methods

Our field research helped us design an online survey of open-data users. The online survey aimed to reach out a wide-ranging open data users beyond commonly cited examples, though focusing on open data use in the UK. Dekkers et al. (2006) noted the difficulty of generating clear sampling frames of open data users. Lacking a clearly bounded population that would allow statistical sampling, the study adopted opportunistic sampling whilst seeking wide dissemination of the survey. A careful balance had to be struck between introducing excessive selection bias by targeting particular open use data communities, and getting adequate responses. A prize-draw incentive was offered to reduce non-response (Couper 2000) and the project blog was developed to demonstrate the authenticity of the research and a commitment of sharing research findings. The survey received 72 responses, 53 provided concrete descriptions of data-use instances, among which 44 from data.gov.uk.

Measures

Following the field research and the literature of open source participation (e.g. Roberts et al. 2006), we conceptualized open data participation as driven by anticipated rewards, use

value and government focus; and the challenges of open-data projects comprising sustainability, social impact and data constraint. We used our field research to identify a battery of items related to motivations, and the perceived challenges and benefits of open-data projects. We then used the Q-method (Stephenson 1953) for sorting items into respective measures of our key variables.

Two researchers sorted the statements independently and compared the classification using Cohen's kappa, and obtained a very satisfactory results (k =.81; 85 percent agreement rate). For motivations, five items were used to describe anticipated rewards (Cronbach's α = .65), e.g. "making a profit", "meeting the requests of a manager/client"; five items for use-value (Cronbach's α = .60), "solving a specific problem", "learning new skills"; and three items for government focus (Cronbach's α = .81), e.g. "making government more efficient, "getting a better understanding of government". For perceived challenges of open data projects, four items were used to measure sustainability (Cronbach's α = .61), e.g. "I will continue to develop and maintain products/sites resulting from this project", "after my experience of this project I am more likely to use open government data on other projects in the future"; three items for social impact (Cronbach's α = .78), e.g. "this project makes a direct impact to people's lives", "this project contributes to the improved local or national democracy"; and three items for data constraints (Cronbach's α = .63), e.g. "the project would have been

different if additional datasets are available", "this project was shaped by the data that was available for it".

For our complementarities measures, participants had to indicate the set of artefacts and technologies that they created for the open-data projects Then we subjected the seventy-two responses to an exploratory factor analysis, and applied the K-1 rule to retain factors with eigenvalue larger than 1. This gave us three factor solutions which were in line with the stack of our complementarities (as shown in Figure 1). We then computed a weighted composite measure, complementarities = 1*SUM ("building a website", "building an API") + 2* SUM ("building a mapping mashup", "creating a visualization) + 4*SUM ("adding to an existing app/website", "building an internal tool for my organization/a client"). The weights represented a very basic form of superadditivity in the relations among complementarities, in this case, a quadratic function (see Tanriverdi and Venkatraman 2005).

Deductive Results

Because of our small sample size, we used path analysis (rather than following structural equations in a simultaneous equation model) to test our hypotheses using AMOS. A covariance matrix of composite measures of all the key variables was used for the path analysis. Figure 1 shows the path analysis based on a combination of several regressions: sustainability was regressed on anticipated rewards (H1); social impact on anticipated rewards (H4) and government focus (H5); data constraint on use value (H3); and lastly,

complementarities on sustainability (H2). The Chi-square shows a good fit of the model to the data and all the hypothesized paths were significant and in the predicted directions.

INSERT FIGURE 2 ABOUT HERE

The deductive findings not only provided an additional testing of the robustness of our inductive findings but also allowed us to test the impacts of motivational drivers on initial and extended participation. The significant and positive path coefficients from anticipated rewards to sustained participation and from sustained participation to complementarities supported Hypotheses 1 and 2, confirming the importance of anticipated rewards in sustaining long-term participation, which is instrumental to accumulation of complementarities. We also fitted paths between other motivational drivers and sustained participation (not shown here), we found no significant coefficients. The significant path from use-value motivation to data constraint (Hypothesis 3) further confirm our prediction that use-value motivation is sufficient to drive hackers to tackle data constraint positively as a challenge rather than a threat.

The findings also indicate support for Hypothesis 4 that hackers who are motivated by anticipated rewards also perceived their involvement in open source project carried a positive impact to the society. Finally, the hackers who were driven to make government efficient and honest were more likely to perceive positively their involvement would exert a wider impact on society.

Discussion and Implications

This study develops new theoretical explanations and presents new empirical evidence to advance our understanding about the conditions for the emergence and sustainability of complementarities of open-data use in the UK. Theoretical explanations and empirical findings of the study make important contributions to a practice-based perspective of open data research and practice. We discuss some of the limitations of the study before discussing the contributions.

The context of inquiry is the open-data use in the UK. The limitation of a single-country study is that its empirical findings may not be generalized to other countries. Yet, in-depth study of open data is necessary for informing future more broad-based studies including cross-national comparison. Also our empirical work has been focused on just one context in which open data can drive service innovation: the use of data by hackers through more-or-less informal hack-day networks and individual engagement with data. Our ethnographic and survey research was designed to identify other forms of engagement with open data, such as by established public service providers or larger enterprises – but at the time of study, no evidence of significant engagement in these contexts was found. It is reasonable to assume, however, that other settings may follow hacker communities in exploring the innovation potential of open data, and such contexts would warrant focus in further research. Nevertheless, the distributed nature of open-data uses makes it difficult to establish the boundary condition of the studied phenomenon. Though not part of our research question, we found a number of active

hackers crisscross and participate in multiple open-data projects. How this boundary spanning activities affect the emergencies and sustainability of complementarities warrants future research.

With these constraints noted, the present study makes several contributions. Our first contribution is to the theoretical base of complementarities. We explicate additional conditions (contextual, intervening, and causal) of complementarities in addition to a contingency account of complementarities which is implicit in the economic version of the theory. We explain why hackers' motivations and hacker culture provides an initial context in which complementarities emerge with the releases of open datasets. This initial motivational context not only drives hackers to expend various efforts (clean, curate, and scrape) but also creates a series of artefact that alters the materiality of technology to facilitate open-data reuse and service innovation.

Our study also contributes to the sociomateriality literature. We show the way artefacts interlock human and material agency provides a basic structure for the stack of complementarities (including combination of external complements) to accumulate. The interlocking sequences explain how complementarities temporarily emerge and are sustained, underlying the significance of motivational drivers and the roles of artefacts to facilitate constant revision and co-configuration between technology and human. This dovetails from prior research on sociomateriality by shifting the emphasis away from resistance and accommodation (as in Pickering's concept of mangle) to a model that takes motivational

drivers and artefacts into accounting for the temporal emergence in practice. The motivational drivers explain why hackers instead of resisting or circumscribing their activities to the limits of open data actively seek to circumvent material constraints by creating new artefacts and inscribing their goals (e.g. wider dissemination) in material agency. The new artefact and material agency serve to scaffold actions and interactions of other hackers (Orlikowski 2006). Both motivational drivers and artefacts are pivotal to address why complementarities are sustained and how they are configured. This leads to our third contribution.

We show not all motivations lead to the sustained participation in open-data projects. Our findings contribute to the literature of open source participation by suggesting that anticipated rewards can sustain longer term involvement, and that the artefacts in form of software codes can induce code reuse which indirectly encourages open-data use. Nevertheless, only a few projects are sustained, and led to the creation of new services. This may be due to the absence of monetary rewards to sustain participation (Roberts et al. 2006), and/or the lacking of an identifiable open-data community to further incentivize and collectivize sustained participation (Fang and Neufeld 2009).

This study also makes a number of contributions to practice. Our framework of interlocking sequences of open data use highlights the various steps involved in moving from the release of data to the development of innovative services. This provides practical guidance for policy programmes seeking to promote uptake of data, and provides an evaluative

framework for looking the value added to data through different initiatives. For example, it would facilitate a comparison of hack-day driven approaches to promoting data-driven innovation, in which a culture of sharing artefacts can support the emergency of complementarities, with competition based approaches to promote data-driven innovation (e.g. Apps for Democracy in the US), in which social and material artefacts may be less readily shared. It also suggests the forms of infrastructure (e.g. source code sharing systems) that open data providers can explore to support more effective use of their data.

Conclusion

Our empirical work shows that there is no straight line from release of open data to service innovation, and that the action of lone hackers is insufficient to realise a revolution in the delivery of services through data. Grand claims for the service revolutions that open data may bring about are overstated; though more modest claims can be grounded in evidence. Our theory of complementarities suggests key conditions involved in the effective use of open data. We find few instances where such complementarities are fully assembled round any specific dataset or thematic area. Drawing evidence from a process-variance methodology, our study explicates the distinction between the conditions for initial participation, and the conditions for sustained participation in service innovation with open data, noting the intricate interplay between human and material agency, mediated through a series of artefacts.

References

Bagozzi, R. P., and Dholakia, U. M. "Open Source Software User Communities: A Study of Participation in Linux User Groups," *Management Science* (39:5), pp. 619-1115.

Berners-Lee, T., 2006. Linked Data - Design Issues. Available at: http://www.w3.org/DesignIssues/LinkedData.html [Accessed May 23, 2010].

Borg, S. 2001. "The Research Journal: A Tool for Promoting and Understanding Researcher Development," *Language Teaching Research*, pp. 156-177.

Brewer, J., and Hunter, A. 2006. Foundations of Multimethod Research, Sage.

Cooper, R., and Haltiwanger, J. 1996. "Evidence on Macroeconomic Complementarities", *The Review of Economics and Statistics* (78:1), pp. 78-93,

Couper, M. P. 2000. "Web Surveys: A Review of Issues and Approaches," *Public Opinion Quarterly* (64:4), pp. 464-494.

Dekkers, M. et al., 2006. MEPSIR - Measuring European Public Sector Information Resources, Available at: http://ec.europa.eu/information_society/policy/psi/actions_eu/policy_actions/mepsir/index_en.htm [Accessed December 21, 2009].

Fang, Y., and Neufeld, D. 2009. "Understanding Sustained Participation in Open Source Software Projects," *Journal of Management Information Systems* (25:4), pp. 9-50.

Fitzgerald, B. 2006. "The Transformation of Open Source Software," *MIS Quarterly* (30:3), pp. 587-598.

Gao, L. S., and Iyer, B. 2006. "Analyzing Complementarities Using Software Stacks for Software Industry Acquisitions," *Journal of Management Information Systems* (23:2), pp. 119-147.

Huang, J., Thornton, K. M., and Efthimaiadis, E. N. 2010. "Conversational Tagging in Twitter," Proceedings Hypertext (HT'10), June 13-16, Toronto, Canada.

Jones, M. R. 1998. "Information Systems and the Double Mangle: Steering a Course between the Scylla of Embedded Structure and the Charybidis of Strong Symmetry," *Proceedings of the IFIP WG 8.2 and 8.6 Joint Working Conference on Information Systems*, Helsinki, Finland, pp. 287-302.

Kuk, G. 2006. "Strategic interaction and knowledge sharing in the KDE developer mailing list", *Management Science* (52:7), pp. 1031-1042.

Lanzara, G. F., and Morner, M. 2003. "The Knowledge Ecology of Open Source Software Projects", *19th EGOS Colloquium*.

Latour, B. 2005. Reassembling the Social: An Introduction to Actor-Network Theory, Oxford: Oxford University Press.

Lee, G. K., and Cole, R. E. 2003. "From a firm-based to a Community-based Model of Knowledge Creation: The Case of Linux Kernel Development," *Organization Science*, (14:6), pp. 633-649.

Leonardi, P. M. 2010. "When Flexible Routines Meet Flexible Technologies: Affordances, Constraint, and the Imbrication of Human and Material agencies," *MIS Quarterly*, forthcoming.

Lerner, J., and Tirole, J. 2002. "Some Simple Economics of Open Source," *Journal of Industrial Economics* (50:2), pp. 197-234.

Markus, L. M., and Silver, M. S. 2008. "A Foundation for the Study of IT Effects: A New look at Desanctis and Poole's Concepts of Structural Features and Spirit," *Journal of the Association for Information Systems* (9:10-11), pp. 609-632.

Mutch, A. 2010. "Technology, Organization and Structure: A Morphogenetic Approach," *Organization Science* (21:2), pp. 507-520.

Orlikowski, W. J. 2006. "Material Knowing: The Scaffolding of Human Knowledgeability," *European Journal of Information Systems* (15), pp. 249-273.

Orlikowski, W. J., and Scott, S. V. 2008. "Sociomateriality: Challenging the Separation of Technology, Work and Organization," *The Academy of Management Annals* (2:1), pp. 433-474.

Pickering, A. 1995. *The Mangle of Practice: Time, Agency, and Science*, Chicago: The University of Chicago.

Roberts, J. A., Hann, I, H., and Slaughter, S. A. 2006. "Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of Apache Projects," *Management Science* (52:7), pp. 984-999.

Robey, D., and Boudreau, M. 1999. "Accounting for the Contradictory Organizational Consequences of Information Technology: Theoretical Directions and Methodological Implications," *Information Systems Research* (10:2), pp. 167-185.

Shah, S. K. 2006. "Motivation, governance, and the viability of hybrid Forms in Open Source Software Development," *Management Science* (52:7), pp. 1000-1014.

Stephenson, W. 1953. *The Study of Behavior*, Chicago: Univer. of Chicago Press.

Suddaby, R. 2006. "From the Editors: What Grounded Theory is Not," *Academy of Management Journal* (49:4), pp. 633-642.

Tanriverdi, H., and Venkatraman, N. 2005. "Knowledge Relatedness and the Performance of Multibusiness Firms," *Strategic Management Journal* (26:2), pp. 97-119.

Volkoff, O., Strong, D. M., and Elmes, M. B. 2007. "Technological Embeddedness and Organizational Change," *Organization Science* (18:5), pp. 832-848.

Von Krogh, G., and von Hippel, E. 2006. "The Promise of Research on Open Source Software," *Management Science* (52:7), pp. 975-983.

Yin, R. K. 2003. *Case Study Research: Design and Method*, 2ed., London: Sage.

Table 1. Open Data Events

Date	Event/Aim
March 2010	A sponsored whole day event, focusing upon the challenges of using location data including Ordnance Survey open data, and user-contributed data. Two hack-days attended by over sixty people, exploring the creation of data and information-driven services to support the core UK government sites.
April 2010	A whole day event, focusing on international development and aid data.
June 2010	The second annual Open Source GIS UK conference (a two-day event), focusing upon practitioners' perspectives on the use and development of geospatial open source tools and geospatial open data; participating in a workshop of the use of Ordnance Survey open data and the Open Space tool for mashup
June 2010	Annual meeting of OSGeo Foundation
June 2010	A sponsored whole day event, focusing upon the UK government open data policy of using public open data to build location-based services
October 2010	The first Open Street Map (GB) meeting (a whole day event), focusing upon the development of a work plan to create a re-engineered version of the Open Street Map database for the GB

Table 2. Processes of open government data use with selected examples

Table 2: Processes of OGD use with selected examples

Process (n=instances)	**Summary** (and example)
Data → Fact *Search* *Browse* ***Extract*** (n=8)	**A dataset is used directly to identify a specific fact of interest.** E.g. Finding out the voting history of a local constituency.
Data → Information *Manipulate* *Statistically analyse* *Visualise* *Contextualise* ***Report*** (n=19)	**Content from a dataset is given a single representation or interpretation that is reported in text or graphics.** E.g. Composing a report that "profile[s] communities of interest within [the local area] as part of the Council's equality & diversity agenda".
Data → Interface *Clean, Combine, Subset Data* *Configure interface tools* *Write custom code* ***Provide interface*** (n=26)	**An interface is provided allowing interactive representation of a dataset – providing information customized to the user's input.** E.g. Creating a searchable interactive online map of stations and former British rail assets.
Data → Data *Convert format* *Filter data* *Augment/combine data* ***Provide API*** (n=17) ***Dataset for download***	**A derivative dataset is provided for download, or access via an API** E.g. I "took Westminster Constituency data, combined it with scraped [General Election] 2005 data and exposed it as RDF."
Data → Service *?* ***Integrate into existing product/service*** ***Create new service*** (n=4)	**A service is provided that relies on open data, whilst not necessarily exposing it to the end-user.** E.g. Using boundary data from the Census to run an application that forwards reports of Potholes to the correct Highways authority.

Table 3. Embedded cases of education data-use

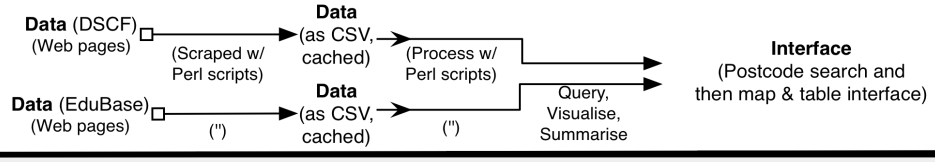
Case ID: Title, *Creator* — ***Sources for embedded case***

Description of data use

Data use schematic

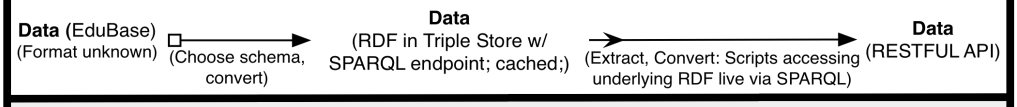
E1: UK Schools Map, Independent developer — **Blog posts, Source code, Direct exploration.**

Searchable list of school locations and performance, presented on a map. Created before widespread availability of OGD in 2006.



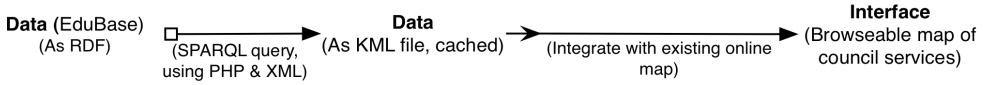
E2: EduBase as RDF, Data.gov.uk team and partners — **Blog posts, Direct exploration.**

Converting the EduBase data to RDF and exposing it via a SPARQL endpoint and API



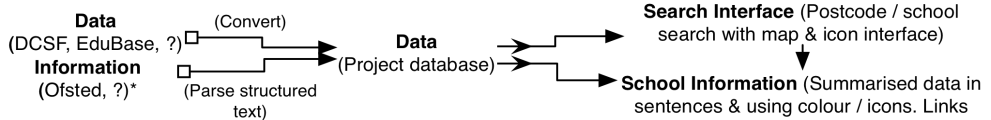
E3: Lichfield My Area Map, Local authority web manager — **Interview, Blog posts, Direct exploration.**

Plotting school locations on a District Council website online map. Sharing code and tools to help others to do the same.



E4: Schooloscope.com, Commercial developers. Funded by 4IP — **Blog posts, Direct exploration.**

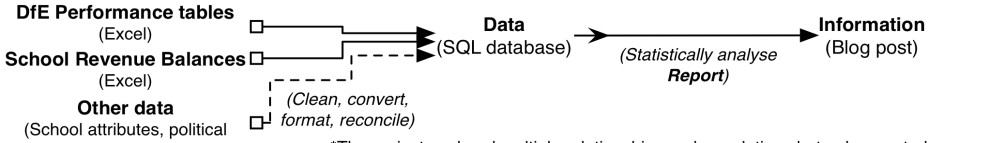
Providing a visually rich interface onto school performance - translating statistics into graphics and plain English statements



**Full technical processes unknown. Certain details are not publicly documented*

E5: School Revenue Balance FAQ, School Governor — **Interview, Blog posts, Survey.**

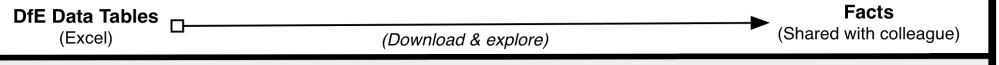
Analyzing school performance against school 'Revenue Balances' and blogging the results.



**The project explored multiple relationships and correlations but only reported on one*

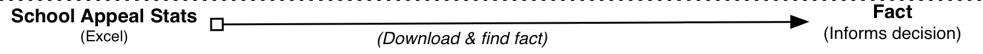
E6: Refugee Charity Research, MSc Student — **Interview, Survey.**

Exploring a spreadsheet of local demographics on young people to inform planning and funding bids for a local charity.



E7: School Admissions Appeals, Project manager — **Interview, Survey.**

Exploring a spreadsheet to identify the success rate of school admission appeals in a specific area and comparing with other appeal success rates.



E8: Guardian Data Blog - Admission Appeals, Data Journalist — **Interview, Direct exploration.**

Datablog entry on school admissions appeals data. Published as Google Spreadsheet; sortable list and heat-map. Written up as news story on the website and in newspaper.

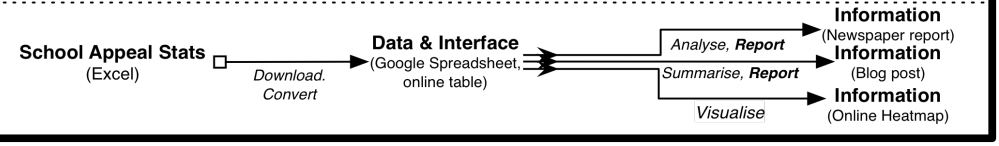
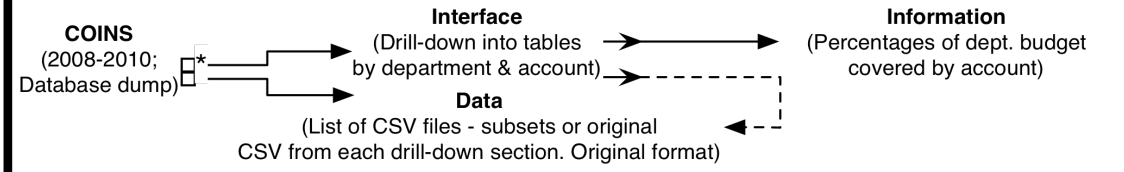


Table 4. Embedded case of COINS data-use

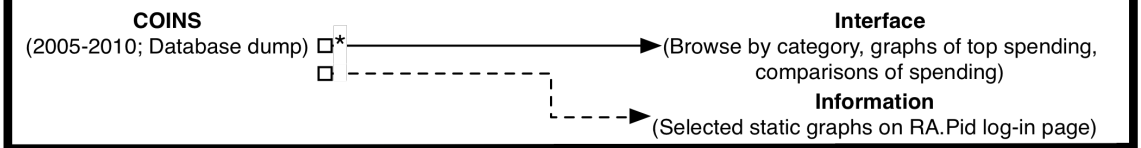
C1: PublicSpendingData.co.uk, Unknown — **Direct exploration.**

Providing interface to browse spending by department or account; and to download CSV files of data subsets. Provides percentages and easily readable numbers (e.g. 1 billion instead of 1,000,000,000)



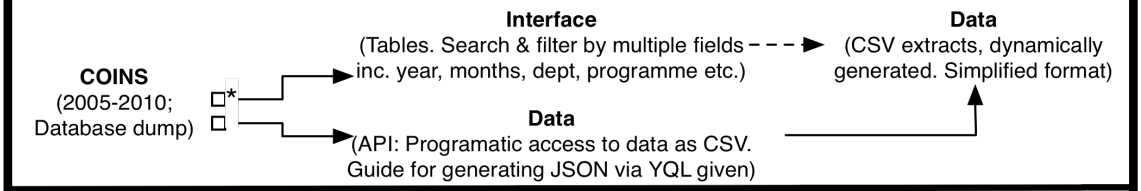
C2: RA.Pid Explorer, Rosslyn Analytics — **Blog posts, Direct exploration.**

Loading the data into an existing spending analysis platform and providing public access to it. Allows navigation and visualization (graphs) of the data.



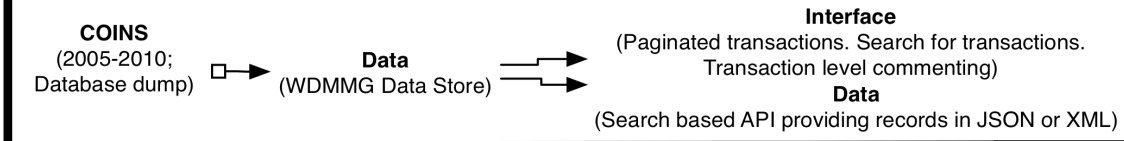
C3: COINS Data Explorer, The Guardian — **Interview, Direct exploration, Documention.**

Faceted browsing interface allowing search and exploration of transactions, with CSV download. Inviting reports of interesting data to The Guardian, who ran a number of newspaper stories based on the data.



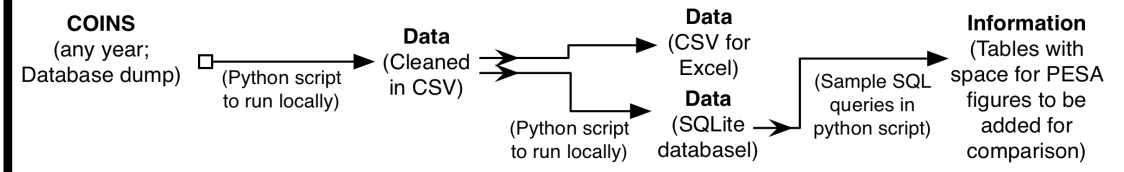
C4: Where Does My Money Go (WDMMG), OKF — **Blog posts, Direct exploration, Documention.**

Loading the data into the WDMMG data store where it can be accessed through a search interface or via a JSON and XML API. Adding comment service to transactions to allow transaction-level commenting. Inviting 'crowd sourced' reports of interesting items to a Google Spreadsheet.



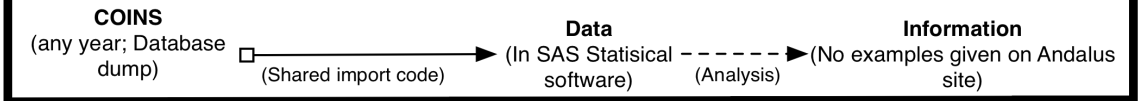
C5: Comparison of COINS and PESA, Blogger — **Blog posts, Source code.**

A set of open source python scripts for processing COINS data to generate reports similar to the official Treasure public spending PESA reports, produced from the data; accompanied by a blog post outlining the process and highlighting differences between the generated and official results.



C6: SAS Import Code, Operational Researchers — **Source code.**

Shared source code for importing COINS data into the SAS Statistical software, posted on the company website.



* Some conversation of the data and storage in a local database is likely to have taken place here.

Figure 1

Assemblage of open-data complementarities

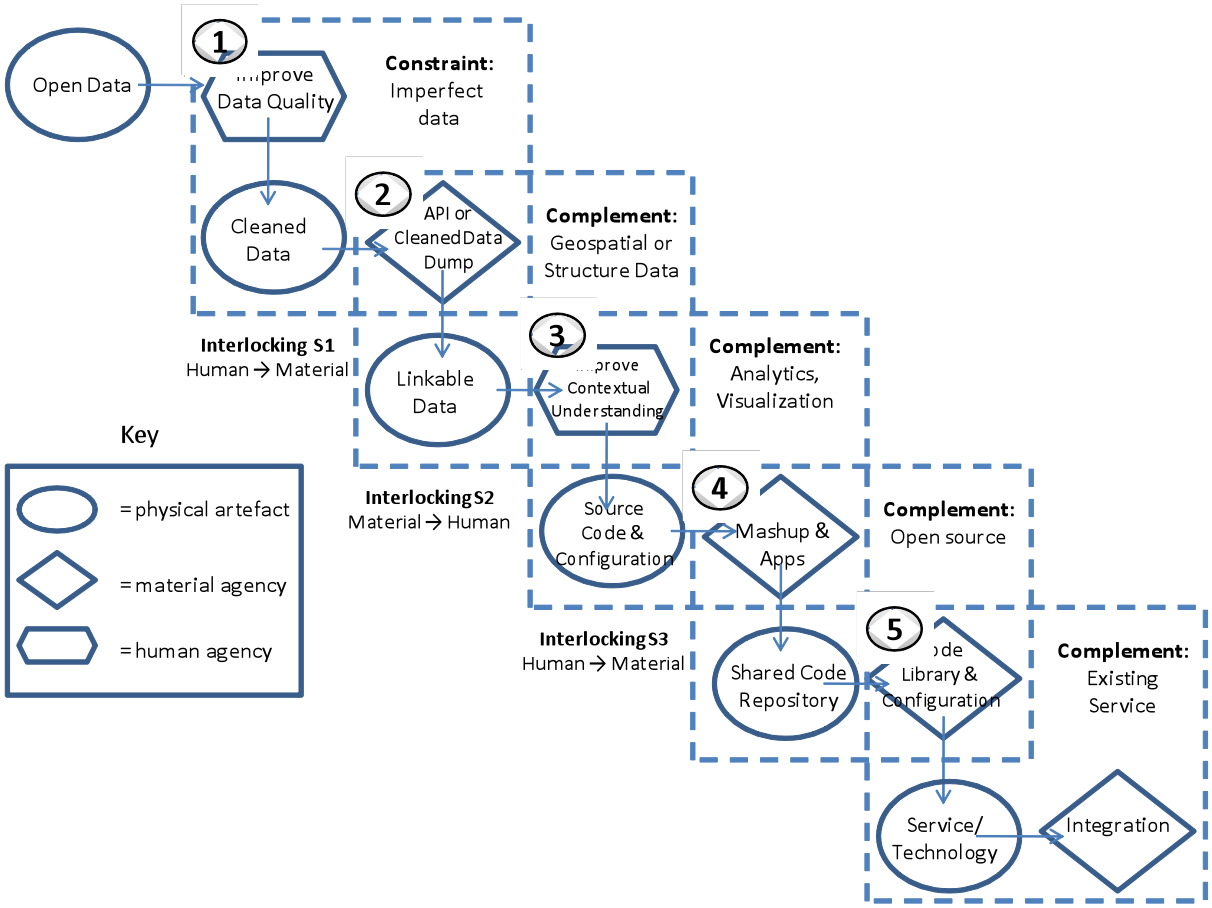


Figure 2

Model Results: Standardized Path Coefficients