ECOSYSTEM INFORMATION MODELS: VISUALIZING COMPLEX DATA TO SUPPORT COLLABORATIVE DECISION MAKING

Dermott McMeel
Department of Architecture, The University of Auckland, New Zealand

Robert Amor
Department of Computer Science, The University of Auckland, New Zealand

ABSTRACT: *There is considerable interest in 'open data' with many administrations launching, or involved in, programmes to make government data open and available. From geographical information systems (GIS) to infrastructure data and building information models (BIMs), it is believed that access to this data will contribute to productivity and efficiency gains. Yet there remains uncertainty surrounding how stakeholders involved in design, construction and maintenance of the built environment might benefit from this unlocked information. We begin this paper by looking at a specific government initiative providing access to built environment datasets; we investigate and compare the different approaches for accessing this information-base. With speculation that open access will lead to huge benefits in productivity, particularly through interoperability, the second part of our paper implements a system to explore the federation of this data and the results of its interoperation in a collaborative visual environment. While prediction models continue to be problematic when simulating multiple complex and interdependent factors of the built environment concurrently, here we appropriate data and exploit it within decision-support systems. A Systems that provides a qualitative virtual 3D rendering of what is otherwise prosaic or opaque technical information, providing the potential to federate, align and compare otherwise disparate sources of data. Arguably access to open data has not revolutionized consumer computing, but it has played an important part in combination with the emergence of other technologies such as mobile devices, Wi-Fi and location aware computing. Here we critique 'open data' initiatives for design and construction, and ask what part they might play—in combination with other technologies—to help deliver on the promise of productivity.*

KEYWORDS: *Design, Digital Media, Interoperability, Data, Ecosystem.*

1. INTRODUCTION

The analysis of *Open* and *Big* data is a burgeoning field of research, where data is made available to the public it is referred to as being open, when there is a lot of data for processing it is referred to as big. Although both have some bearing on this research we concentrate, in this paper, on the implications of open rather than big data. It is useful to consider for a moment the variety of drivers, interpretations and intentions that inform 'open data' and it's manifold of meanings. Within the context of this paper it is worth reflecting on open data in relation to *governance* and *open source*. The former having its roots in the freedom to access information through the likes of the freedom of information act (FoI), and the latter largely informed by freedom to use, share and modify some digital *goods*. While these are not necessarily oppositional intentions, they are clearly ideologically different; the applications developed through this research as described later in section 3, illustrate the impact these different ideologies have on the extent to which data can be shared and used. The term *open data* is used very generically and often woven into promises to improve productivity and value. In New Zealand the Productivity Partnership and GeoBuild strategy are bringing the subject of open data into sharp relief for architecture, engineering and construction (AEC), where a 20% productivity improvement for the AEC sector is targeted for 2020. Yet already there is skepticism emerging within the field and some nascent work suggesting gaps between the promises and benefits of open data (Janssen, Charalabidis & Zuiderwijk 2012). *Productivity* is also a generic term; analysis has revealed dozens of disparate factors contributing to notions of productivity in the AEC sector (Borcherding, Palmeter & Jansma 1986). In fact as the number of notable projects outlined in Table 1 attest, there are unique measures of productivity appropriate within different situations. The trend to improve productivity continues with the formation of the UK BIM Task Group, who ring-fence *reduction to capital cost* and *carbon burden* as their metrics; interestingly they have almost completely dispensed with using the word *productivity*. Continuing to reflect on Table 1 we might also note the focus is almost exclusively on national and multi-national companies undertaking projects of significant

Citation: McMeel, D. & Amor, R. (2013). Ecosystem information models: visualising complex data to support collaborative decision making. In: N. Dawood and M. Kassem (Eds.), Proceedings of the 13th International Conference on Construction Applications of Virtual Reality, 30-31 October 2013, London, UK.

scale. Countries like New Zealand, for example, have considerably different market forces to Europe or America, it should not be assumed the cost/benefit ratios from one country can, or should, be automatically applicable to other geographical or economic contexts.

Table 1: Projects citing productivity improvements (reproduced with permission of V. Gonzalez).

Country	Project Name	Project Type	Type of Improvement			
			Cost	Time	Productivity	Others
U.K	Barts And The London	Commercial multi project	22% cost-benefit	6 weeks saved	Clear deliverables & procedures	Improved safety for FM and Users
Denmark	New HQ for Engineering consultant Ramboll	Office Building	Direct benefit DKr 3.8 million	Saving in time for zero RFI		
India	3x660 MW Supercritical Thermal Power Plant	Infrastructure project	Under budget	Reduce engineering time by 10 percent		Multiple design simulation
Brazil	Matec Engineering	Construction Project	Enormous saving	Saving 40 percent in time	30% material saving	
Finland	Aurora 2	Education facility at Joensuu University	Under Budget	Saving design change time	Energy efficient	Sustainable features

1.1 What is Open Data?

The general supposition of open data is that information generated in one place is useful in others. Currently access might be restricted, prohibited or laborious, and by making it open individual stakeholders can increase speed, innovation or enable them to do more with less. Government and building industry partnerships are thus approaching the concept of *open data* with much in common with the open source movement, rather than traditional government freedom of information (FoI) acts, and we feel there are important distinctions here.

Open data is based on the presumption numerous people will access, modify and redistribute some digital goods, which can cause fragmentation; a phenomenon whereby multiple versions of a *thing* cause some versions of the thing to be incompatible with others. For example phones running the Android operating system have been criticized for suffering from fragmentation, where software applications run inconsistently across different models of handset. Thus one of the main challenges with open data is tracking access, problems, reliability, changes and version control. The open source movement has developed numerous version control systems to do this, such as Git and Subversion. GitHub (https://github.com/) is an online version of Git that assists collaboration and project management, it allows complex code to be changed and tracked in relation to how it works—or no longer works—with other pieces of the project which are also changing concurrently. Git has even been appropriated as an issue tracker for house maintenance (McMillan 2013) and elsewhere the authors are exploring if it has merit as a tool for tracking the data and its changes during building design and construction. Essentially the open source data ideology has evolved systems designed to track the changes and relationships within and between complex dynamic datasets. Open data in relation to governance is somewhat different with its roots generally found in various country specific FoI acts, which enable access to specific static and usually historic documents. Accessing information through FoI can be laborious, although open government data is a notable improvement on FoI it has emerged from a culture of simply providing access to specific information. What will be revealed in this paper through the development and analysis of two software applications drawing on open data is the different cultures that underpin the provision of data are highly influential; they ultimately have considerable bearing on what can be meaningfully achieved.

Elsewhere the authors have explore the impact of the materiality of data on design and making processes (McMeel & Amor 2013) and our findings here continue to support the suppositions put forward by Paul Dourish that data storage and rendering are not incidental, they are fundamental and influence understanding and knowledge practices that surround them (Dourish & Mazmanian 2012).

1.2 Two Software Applications

An initial brainstorming session attended by designers, an architect and computer scientist revealed various personal frustrations with professional and consumer activities. How they are laborious and time consuming, or even in light of current services and data available online they remained unchanged for decades. Two of these scenarios, one oriented towards professional designers and one directed at house buying consumers, were chosen as test cases because of their potential to scrutinize suppositions regarding open data within the AEC industry. The first scenario explores how to improve the process of site analysis, which is a specific activity undertaken by architects at the early stages of building design. The second scenario investigates improving how prospective house buyers can evaluate the suitability of prospective house locations.

In a break from conventional methodology we did not establish a brief or set of functions and interaction capabilities; we instead chose to outline a narrative for both the professional and consumer scenario. Our decision was in part informed by the sixth sense transport research project (http://www.sixthsensetransport.com/) that uses similar approaches to successfully keep projects and research people-centered. Where research is overtly technical in nature, it is all too easy to lose sight of the goal and instead focus only on the development of technology. The scenarios provided a qualitative rather than a quantitative framework within which the designers considered choices and decisions. Through this work we aim to advance generic discourse and supposition that surrounds the provision of open data to increase productivity. By identifying and focusing on these two scenarios we look at explicit situations and how they might benefit through access to available digital information. The scenario for the professional application was that an architect prior to visiting site could—through this software application—access data that is usually time consuming to gather and aggregate. Such as topology, soil type and building usage; this data could be rendered to build an overview of the site. The scenario for the consumer application was software that a house buyer could use to identify prospective house locations and be quickly presented with information that enables evaluation and comparison. In the following sections we will outline the data that was available at the time of writing, and then explain the software implementations before finally answering the following questions. Are current open data initiatives adequate? What part might they play in improving productivity in the AEC sectors?

2. A TYPOLOGY OF DATA

There is an ever-increasing variety of data related to the built environment. As well as Google Earth and Maps, there are localized providers of amenity databases and service such as Yellow.co.nz, Zenbu.co.nz as well as government initiatives. Whereas commercial providers like Google invest heavily in the rendering of data to make it useful, visualization of data is not the core business of government data providers. Consequently government occasionally comes under criticism for providing opaque and unintelligible data (Smith 2012). Yet it can be highly revealing and valuable, so in this section we will consider the data available under three categories, *government silo'd, government managed* and *commercially provided*.

2.1 Government Silo'd Data

The word *silo* refers to the storage structure of data, one that requires traveling deep into a silo to access data and to leave one silo entirely to journey deep into another. This leaves data mining difficult and time consuming. Even where governments provide dedicated data sites such as (https://data.govt.nz/) it can be difficult to find what you are looking for. With a data silo a number of generalizations can be made. First, the data is typically available in a number of file formats such as excel spreadsheet (XLS), comma separated variable files (CSV) or a portable document format (PDF). Second, it is unclear if said data is stored in databases and automatically exported in the requested format or previously saved in these formats and stored on a server until requested. Third, in these instances data usually requires manual *point and click* to download the data to your computer hard drive. Data in PDF format is difficult to mine with computers; XLS and CSV are reasonably common file standard for data and are machine readable. CSV was the format utilized most often in this research, although as we shall see during our analysis it was not without problems.

Relatively speaking, because it is not machine readable, a government data silo presents obstacles for data access when compared with commercially provided systems that use application programming interfaces (APIs). Access requires manual intervention by a person to physically *point and click*; this limits utilization by automated and interoperable data systems. Yet governments hold valuable statistical information on, for example, crime, mortality and changes to land use; data that is useful in grasping historical trends, current state of, and potential future directions for any given area. It helps to grasp a location's flavor or to use the German term *gestalt*, which Oxford English Dictionary defines as a '*shape, configuration, or structure which as an object of perception forms a specific whole or unity incapable of expression simply in terms of its parts'*. The point being the cognitive impact of a particular combination of datasets—or overview—can be qualitatively different to analyzing them in isolation. What is often called the 'overview effect' (White 1987), was originally coined in referring to the impact on astronauts of seeing the earth from space:

The Overview Effect is the experience of seeing the Earth from a distance, especially from orbit or the Moon, and realizing the inherent unity and oneness of everything on the planet. The Effect represents a shift in perception wherein the viewer moves from identification with parts of the Earth to identification with the whole system. (White 1987, p. 38)

While the Overview Effect is best understood through narratives from astronauts, it is also useful as a way to frame the problem of our data rich world, where it is increasingly difficult to get an overview when overwhelmed by granular and detailed information. This position has been advance by Chris Speed who suggests our data intensive environments produce an *Underview Effect*, which provides a greater awareness of social and geographical context (Speed 2010). Both effects are valuable within the context of our scenarios, however it is the Overview Effect that is harder to achieve through data, as data—according to Speed—is predisposed to producing an Underview.

2.2 Government managed

Moving on from government silo'd data we found several sources of information best described as government managed. Where a geo-data management service is used to organize and facilitate access to information. In New Zealand, government departments such as Land Information New Zealand-LINZ (http://data.linz.govt.nz/) and research institutes like Landcare Research Information System (http://lris.scinfo.org.nz/) have deployed a system developed locally by Koordinates (http://koordinates.com/); an award winning system (Sweeney 2012) designed specifically to manage geo-data.

While the core functionality of the Koordinates system is data management rather than visualization, however the inclusion of a map makes understanding the geography of the data much easier. Still, with over one thousand datasets available on LINZ, knowing which to access continues to be challenging. A distinct advantage of the Koordinates system is that it provides an API, enabling data access and manipulation without the need for manual download. This creates the possibility of developing software and mobile applications that can communicate with, but created independently from, the data; as long as the data management is kept up-to-date, the application will be up to date. The data is always presented in a consistent format in this case it can be called as extensible markup language (XML) or JavaScript Object Notation (JSON), both of which are international and widely used web standards. This consistency means manipulating the managed data is much easier and, we found, provides greater flexibly than silo'd data.

2.3 Commercial Data

The final category of data is *commercially provided*; this data usually provides information on local amenities such as bars, restaurants, shops etc. It is made available by commercial groups such as Google, Localist.co.nz or Zenbu.co.nz, often gathered using a variety of collaborative or crowd sourced techniques the data also changes more frequently than government data. It is highly sought after as it gives feedback on currently available amenities. Much like the Koordinates systems discussed in the previous section, these providers also typically provide powerful APIs to enable sophisticated manipulation of their datasets.

Returning to our test case scenarios, both developers agreed it was necessary to draw on the commercially provided data provided by Zenbu.co.nz, however the developer of the house-buyers application would use Koordinates and LINZ managed data while the developer of the professional site analysis application used government silo'd data, in the following section we will compare and contrast these two applications.

3. SOFTWARE IMPLEMENTATIONS

The two scenarios and software applications being discussed in this section are technically quite similar. Both visualize geo-data, in some cases the same geo-data, in a virtual environment to assist individuals or groups of people making decisions or judgments about specific places. The difference between the two is best articulated in terms of White's *Overview Effect* and Speed's *Underview Effect* discussed earlier in this paper. The consumer application targeting house buyers will need to render this granular data qualitatively providing the user with an overview to make quick comparisons between locations. The professional application directed at professional architects will need to present the *underview* that we have just discussed and the detail of amenity and geographical context for it to be of any use within an architectural site analysis. Both developers chose to use the Processing software (http://www.processing.org/) to develop their prototype applications.

3.1 Professional site analysis application

Without wishing to overstate the obvious, the aim of a site analysis is to gather pertinent information on a place that will help inform the design of a building intended for that specific place. With the exception of procuring information through geographical information systems (GIS), the process of gathering data for a site analysis is much the same as it was twenty years ago; a large portion of relevant data (foliage, building typology, traffic/pedestrian flow, local amenities) is gathered through visits to site. Although much of this data is available online, the aim of the professional application was to allow a user to focus on a particular location and render pertinent data. There was no presumption this application would replace a physical site visit, rather, could accessing and visualizing soil, contours, land use and amenities in this way be more efficient?

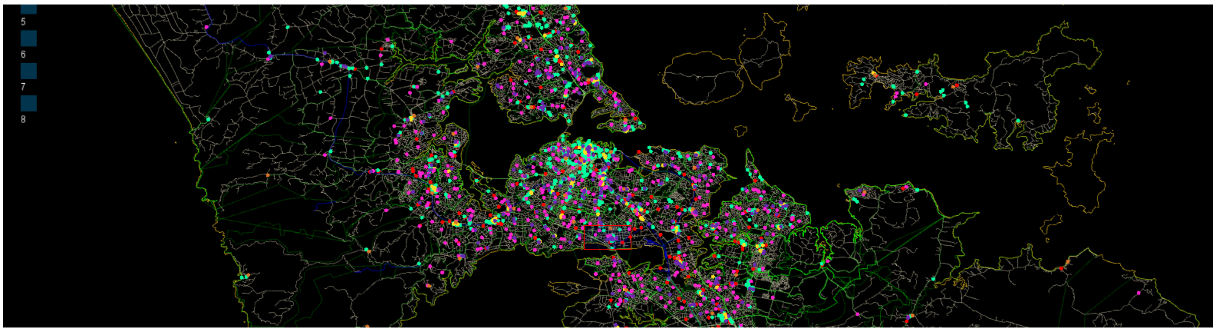

Fig. 1: Government silo's data and LINZ data rendered from downloaded CSV files.

Initially obtaining the data was relatively straightforward, data was downloaded as CSV files that would be imported into the application. Even at this early stage a number of problems began to emerge, beginning with legibility and data integrity. Each downloaded data set was quite different with little consistency in format and data identifiers were overtly cryptic. For example the column header for the land use data set were mostly meaningless without further explanation: OBJECTID, TYPE, LEGEND, LUC1C, LUC1S, LUC1UU, LUCPL, LUC2C, LUC2S, LUC2UU, LCORRC, LCORRS, LCORRU U, LCORRQ, LUC, LCORR, WKT. When this was deciphered the data was still returning an error within the application because of particular encodings used for some strings of characters within the CSV file. These had to be manually replaced within text editing software before the data could be used reliably for a virtual rendering of the city.

Having made the data machine readable a number of issues continued to hinder progress on the professional application. In certain data sets urban parks were marked as single GPS points rather than a bounding polygon that would helpfully represent the geographical extent of the park. Also the road networks were stored as vector lines and in the absence of contour data – which was not free but would require a fee to download. The developer invested considerable time to return the rendering of the data in Figure 1. In this rendering land usage is color coded and a key with checkboxes is being implemented on the left hand side of the image to enable layers of data to be turned on and off.

In summary, although government initiatives are making much data available, where it is provided in these governments silo's (XLS, CSV) is was not directly machine readable; it required manual downloading and editing. The professional application, although functional, did not progress as far as was anticipated because of the numerous data irregularities and inconsistencies. It was also necessary to hard code the data into the application, which will result in it becoming obsolete as the data becomes outdated unless the application is manually updated.

3.2 Consumer house buying application

Turning to the house-buyers application, one of the authors purchased a house just prior to this research. The Auckland property market is highly competitive, by way of an example it is not unusual for houses to be sold within one week of listing and before any public viewing of the dwelling has been possible. In such a marketplace it is difficult to gather enough information and get to know a neighborhood within a useful timeframe before committing to a purchase. With this scenario we knew the relevant data was available online, the aim of the house buying application was to give a user a visualization of a suite of data sets that would otherwise take considerable time and research to obtain. It was an attempt to provide a rendering of a qualitative overview, the gestalt, of highly granular and detailed information. The strategy adopted by the developer of this application was different to the previous professional application; it was decided in this case to use API's where possible. Although during development some historical seismic data from GeoNet was hard coded to show a chronology of seismic activity across the country in combination with the other data. In the finished application (Figure 2) there was virtually no hard coded data present, it was all drawn dynamically when required from Google Maps, Koordinates, LINZ and LRIS.

Fig. 2: House buying application showing closest amenities, supermarkets and school zones for two locations.

The methods for data appropriation used in this application were considerably superior to those used in the first application, although the data required for the first application was not necessarily available through one of the API enabled sources. This application only requested data from data sets when a location was selected on screen, thus it was highly efficient and fast. It also remains current, the amenity data is drawn from Zenbu.co.nz and thus continues to be updated and relevant.

4. CONCLUSIONS

Let us return to our opening question – what part might open data play in the promise of productivity? Firstly *open data* and *productivity* have been found to be overtly generic and ill-defined terms. Our opening reference to Borcherding analysis of *productivity* shows dozens of factors that potentially contribute to productivity. What is required initially is specificity, the UK BIM Task Group has, for example, ring-fenced capital cost and carbon burden to be the national metric for the purposes of focusing actions, research and initiatives. Although there might be disagreement over the relevance and impact of these factors, we can agree that providing such a focus, for such a complex and varied industry, is necessary to mobilize key stake-holders and, returning to the theme of this paper, identify what the information and inter-operability requirements are, and in what roles do they need to be implemented. Although we found an abundance of information, it did not always meet the inter-operability needs of our scenarios. Scenarios, which are modest in comparison to the mobilization that will be required for government initiatives.

Turning to our two understandings of open as being either (1) Informed by the culture of Freedom of Information (FoI) or (2) Informed by the open source movement, the future of design and construction will not be transformed by open data driven by the culture of Freedom of Information. Data from silo'd government repositories required manual downloading and usually manual editing. In the race to provided easy to use tools and services for the industry, this is highly problematic. Where our scenarios benefited substantially from the provision of data it was informed by the open source movement. A modest easy to use consumer application was designed and deployed that continues to return relevant and current information as it accesses the reliable and current data provided by Koordinates, LINZ and Zenbu through their API's. It has proven so successful in fact that, at the time of writing, the developers are exploring options for commercialization of the consumer application where this information can be accessed through a mobile device. All of this is possible as there is no inherent obstacle within the data or its access that impedes these entrepreneurial initiatives. Our unconventional approach to focus on two scenarios proved highly valuable. It prevented the technology taking center-stage and helped maintain a focus on the practices and the needs of the users of the technology. One issue that remains unresolved surrounds the problem of rendering *fuzzy* data. Some of the highly critical government data, such as school zones and school decile rating (a New Zealand method for ranking school performance) changes occasionally. When data is rendered it becomes fixed and yet it is important to convey that some information is fluid and subject to change, particularly highly valued information.

In summary there remains much hyperbole around open data and productivity, what is clear from this modest programme of research is that simply encouraging or mandating that information is provided will not necessarily guarantee the data access or inter-operability that researcher, innovators and companies will need to meet the approaching industry challenges. Although the problems cannot be completely predicted and their answers not known, we will be more likely able to deal with them given access to open data as provided by Koordiates, LINZ and Zenbu. In fact at the time of writing the US government has mandated that government data will not only be open but *machine readable*. While our test cases are admittedly modest in scope the developer of the consumer application is investigating commercial development. Perhaps the most pertinent point is when data is opened through powerful and flexible APIs and managed by people or organizations with key competencies in the management of geo-data it becomes possible to conduct research and develop services, devices and applications that address these challenges. It becomes possible to identify problems and develop solutions, and it would appear to be the means by which we will advance the industry rather than the obstacle that impedes it.

5. ACKNOWLEDGEMENTS

The authors would like to thank Paul Organ and Kurt Meyer who developed the two software applications during the University of Auckland's Summer Scholarship programme, as well as Hamish Campbell from Koordinates.

6. REFERENCES

Borcherding, J. D., Palmeter, S. B. and Jansma, G. L. (1986). Work force management programs for increased productivity and quality work, *EEI Construction Committee Spring Meeting,* 4–9.

Dourish, P. and Mazmanian, M. (2012). Media as Material: Information Representations as Material Foundations of Organizational Practice, In P. Carlile, D. Nicolini, A. Langley, & H. Tsoukas (Eds.), *Perspectives on Process Organization Studies Vol 3: How Matter Matters: Objects, Artifacts and Materiality in Organization Studies*, Oxford: Oxford University Press. Retrieved from http://www.ics.uci.edu/~mmazmani/Site/Publications_files/dourish_mazmanian_materiality_final_book_submission.pdf

Janssen, M., Charalabidis, Y. and Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, Vol. 29, No. 4, 258–268. doi:10.1080/10580530.2012.716740

McMeel, D. and Amor, R. (2013). Fabricate It, Paint It – And Don't Wait up: Separating Fact from Fiction in Digitally Sponsored Fabrication. In J. Zhang & C. Sun (Eds.), *CAAD Futures,* Berlin: Springer Verlag, 149–158.

McMillan, R. (2013). This Old (Open Source) House: Man Renovates Home on GitHub, *Wired,* Retrieved April 17, 2013, from http://www.wired.com/wiredenterprise/2013/01/this-old-house/

Smith, M. (2012). The big debate: open data, *The Guardian,* Retrieved April 17, 2013, from http://www.guardian.co.uk/public-leaders-network/2012/aug/20/big-debate-open-data

Speed, C. (2010). Developing a Sense of Place with Locative Media: An "Underview Effect", *Leonardo*, Vol. 43, No. 2, 169–174. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/leon.2010.43.2.169

Sweeney, S. (2012). NZ Gov wins prestigious spatial award, *FutureGov*, Retrieved April 17, 2013, from http://www.futuregov.asia/articles/2012/nov/23/nz-gov-wins-prestigious-geospatial-award/

White, F. (1987). The Overview Effect, Boston: Houghton Mifflin Company.