Open Research Online

The Open University's repository of research publications
and other research outputs

Creating a Datascape: a game to support communities in using open data

Conference or Workshop Item

oro.open.ac.uk

Creating a Datascape: a game to support communities in using open data

Short Paper

Annika Wolff
The Open University
Milton Keynes, UK
Annika.wolff@open.ac.uk

Matthew Barker
The Open University
Milton Keynes, UK
matthew.barker@open.ac.uk

Marian Petre
The Open University
Milton Keynes, UK
m.petre@open.ac.uk

ABSTRACT

Open data is a resource that both local governments and citizens can use for answering questions about a particular community and for innovating new products and services. In reality, there are a number of barriers to the use of open data, not least that citizens can find it difficult to make sense of available data and understand what sort of problems it may solve. In this paper we describe the development of a board game that is designed to help people to better understand the relationship between data, the environment from which it derives and the questions it can be used to answer. We describe the motivation for the game's development, its current design and our ultimate intention, to create a tangible interactive, datascape. This work has been undertaken as part of a smart city educational programme.

CCS CONCEPTS

• **Human centered computing** → Human computer Interaction (HCI) → Interaction paradigms

KEYWORDS

Board game, map, open data, tangible, communities, smart city

INTRODUCTION

Increasing quantities of data are being collected from the environment and the people who live and work in it. For example, data is collected from environmental sensors, web behaviour, shopping transactions, personal trackers and surveys. A vision of an instrumented, interconnected and intelligent Smart City described by IBM [5] portrays data as driving innovations that improve the efficiency and sustainability of cities, providing solutions to the everyday problems that citizens face.

Smart cities often aggregate data into single data hubs. This process is intended to support both innovation and the creation of products and services by making data easier to find, query, combine and use. There has been a recent push to open up the data in such repositories for use by the public. This means that it is not just city officials, or technology specialists, who can use the data to identify issues and to create new technologies, but the public

themselves. Open data is defined by The Open Data Institute as data which 'anyone can access, use or share' [8].

The reality is, however, that the public are under-utilizing the available data as a resource for question-answering and design. Janssen et al. [7] performed interviews with members of the public to gather evidence of their opinions and experiences with open data. His findings suggest a number of barriers that currently prevent citizens from converting open data into something that they can use. Key among these is that citizens lack the knowledge to make use of, or to make sense of data.

Our objective is to explore whether connecting a tangible map-based interface to a data hub (as an alternative to more traditional types of interface) can help users to make better sense of data. Conventional databases lack contextual information and don't show where the data physically came from. Data is normally presented in a table and accessed via filters. Our idea is to use a map as the filter and allow users to access data by location. We have coined the term Datascape (changed from previous term DataPlace) in contrast to database, to convey how the data and interface would function. The term database implies that the data is static and stored in a specific location, i.e., in the computer. The term Datascape is intended to convey that data is in fact gathered in real time from physically distributed sources.

We have prototyped a board game for querying data. We chose a game as it helps us to meet the dual objectives of our study. We are working as part of a smart city project team assessing new ways of educating people about data in smart cities. Designing a game was an opportunity to test the Datascape concept and at the same time formulate a new approach to educating people about data in the city. The board game provides a way to engage people with data quickly and allows them to explore ideas within a strict game narrative. The game rules limit the ways they can interact with the data and allow us to test the Datascape concept in a controlled way. This allows us to focus on specific research questions, such as to what extent the interface supports users in selecting appropriate data to answer questions. The game is being developed in two stages. Firstly, there is the simple, non-digital, board game version. The next stage will be to embed 'live data' into the board game using Near Field Communication (NFC) tags so users can receive live data by querying different areas of the map with a mobile device. The remainder of this paper discusses

the background and evolution of the game, from paper prototype, to functional board game and finally to designing the interaction.

BACKGROUND

Jansen et al. [6] define advantages of physicalized data over other forms of representation: *leveraging perceptual exploration skills*, *making data more accessible and engaging for a broader population,* and *facilitating learning and understanding.* Based on these ideas, a number of tangible tabletop interfaces have been developed to allow analysis of large amounts of data. One example is inFORM, which uses programmable rods to give data a physical form [2]. Such interfaces encourage group discussion and collaboration, but have some drawbacks in that they are large (and so difficult to move around) and expensive. In contrast, other studies have focused on providing interfaces for contributing or viewing data within the natural environment. Goljstein et al [4] designed the Voxbox, a physical questionnaire system. This is an example of a tangible data collection device with a series of curious-looking levers and buttons that make it more attractive and engaging than conventional paper questionnaires. The Datacatcher, [3], meanwhile, aims to prompt curiosity about data through a hand-held device that can be carried around and which changes to display data that is relevant to the users' current physical location, thus strengthening their understanding of the relationship between the data and the environment. However, the time spent in travelling through the physical environment may make it hard for people to recognise relationships between data sources that are not close to one another. This was addressed in the DataCatcher trial by trying to use the device over a short time period, a number of days.

The Datascape is designed to overcome this limitation by creating a 'bird's eye' view of the world through which data from a large physical environment can be queried in a short time period, without the need for travel. The Datascape is initially presented as a game that is intended to be used with communities who are starting to think about exploring open data. These might be communities trying to use data to identify or verify problems, e.g. with local services or citizens who are building new applications from the data to serve a particular purpose. The game is being developed through several phases, now described.

GAME DEVELOPMENT

The aim of the first prototyping stage was to discover the connections that people might make between questions and different types of data and how they would relate this to a map. Specifically, we wanted to know a) what would be the important features of a map interface that would also be the 'board' for the board game and b) what data types should be included. The first prototype consisted of a map printed onto A3 paper, which represented a fictional place and contained a number of different features, including: green areas; water; roads; hills; residential areas, school; church; shops, factory; farmland; business area. There was a set of data markers, each of which presented a different data type (e.g. wind sensor data, or survey data) that could be placed onto the map to represent a point of data

collection. We started out with three questions: 'Where would you site a windfarm?', 'where would you site a nature trail?' and 'what renewable energy sources would be most suitable for this fictional town?' The participants were a range of people with some interest, or in some cases a great deal of expertise, in smart city technologies who were attending a smart city project event. This version was not framed as a game, instead the aim was to capture the collective expertise of participants to guide the initial game design. Participants worked in groups, of mixed expertise, to discuss where to place data markers that would 'collect the data' that they thought would have the most value for answering each question. There were 5 groups (of 5-6 people in each group) and they answered up to two questions each. Tracing paper over the map allowed participants to write ideas and comments. They were also allowed to propose the creation of new data types.

Information collected in the workshop was used to choose an initial set of 16 data types for the game, based on the types of data the groups had used or proposed. These were *light*, *weather*, *wind*, *water level*, *temperature*, *air quality*, *busyness* and *noise sensors*, *recycling points*, *twitter feeds*, *photographs*, *dialogue data*, *maps of solar panel coverage*, *crime data, soil moisture* and *traffic count*. The map itself was found to be very effective at prompting discussion in the group about different data types and the questions. Therefore, it was decided that no changes would be made to the map. The three questions trialled prompted a lot of interesting discussions and so were kept for the next stage of prototyping.

Mock Up And Testing Of Game Mechanics

Figure 1. The map ©Yen Quach

Drawing on the findings from this session, a first version of the game was then created. In the game, the map is divided into a series of squares (figure 1). Each square is associated with between 1-3 data types. Each source of data on the map was allocated a score to reflect how useful it would be for answering a particular question. This score was decided based on a) analyzing where experts in the first prototyping stage had placed markers b) the knowledge of the game designers about data. The players' goal was to 'find' the data that answered the question (unlike the first phase where people could choose where to collect the data). Players would choose a square on the map to query and then would look at a game card for that square that would show what

data was available and the scores. A card is shown in figure 2 (left-hand side).

The idea was to simulate a real life data hub, where people might want to locate data about a particular location, but still find that the data had not yet been collected (something that is quite a common experience when using real data repositories). Players would pick six random data markers - the data marker set contained 3 of each data type - then they would take turns to query a square and decide if they wanted to 'claim' the data to help them to answer the question. Some sample markers are shown in the right-hand side of figure 2.

Figure 2. A card showing data available for a map square and some data markers

The goal of the game was to place all markers on the map and for these to total the highest scoring datasets for the current question. An element of chance was introduced through cards that the player took on each turn, which might provide extra scores for players combining different types of data in a particular location, or collecting the same data from different parts of the landscape (e.g. placing wind sensors on 3 different squares). The game was designed so that the learning goals that motivated the creation of the game were aligned with the player goals for success in the game. The *learning goal* being to help the player to understand different types of data that can be collected, where it would typically be located and how it could be used to provide answers to questions. The players' goal being to score the most points, by collecting the 'best' data. To do this, a player would have to be strategic and try to get data markers that reflect the most useful data sources and be the first to query places on the map that are most likely to have high scoring data of that type. The game included the possibility to swap data during the game, through picking a suitable chance card.

This version was made into a fully working board game, which was trialed with a group of board game enthusiasts and then also used at a public engagement event in a museum, with a slightly simplified scoring approach suitable for children (see figure 3). In the trial with the board game club, players were asked to play in pairs so that it was possible to capture their strategic thinking and to make it easier to identify problems in the game play as they were encountered. Feedback identified where elements of the game play did not work very well. For example, in this version players could place a sensor and there would be no data or 'score' in that location. This was because this version of the game was presented as trying to find data in a data hub, where data for a given location may not exist. Interestingly, while this was

problematic for the older 'game players' it was considered a fun feature for younger players who enjoyed the random chance.

Figure 3. Trialling the board game with a young audience

Based on the overall feedback from the game club session, the final version of the game is currently being designed. In this version, players are once again given the role of deciding where to start collecting data from. This means that each square on the map has to be a viable place for data collection, but that a) some squares will yield better data than others, based on geographical features (e.g. wind data is not very good from residential area and survey data not good where it is not densely populated and there are fewer people) b) some data will be more appropriate for certain questions than others (for example, wind sensor data is more appropriate to answer a question about where to site a wind farm than air quality data). The number of data types is being reduced to between 7 and 10 (still being decided), to make this more practical. A much more rapid fire game mechanism has also been chosen. In the previous version, players would query, decide and score on each turn. This could make games last a long time for a single question. In the final version, players must still take turns to place markers, but they will place them all at once before scoring both teams for each question, and there will be a much larger number of questions. In this way, they have to be more strategic over where they place each data marker. There are two parts to the score. There is a score for placing a data marker in a 'good' location, for example placing wind sensors in a windy spot. This is fixed for all questions, so players should be able to improve over time as they learn (this will be evaluated during trials). Next, there is a score for placing data that is associated with a particular question. This is written on the back of each question card (that the player does not see until they start scoring) and gives bonuses for either placing a good combination of different data types on a single square, or the same data type (that is relevant to the question) across multiple squares. For example, a player may get a bonus for placing three wind sensors in answering a question about the wind farm. This is similar to the previous version, but presented slightly differently. This bonus is multiplied by the score for the relevant data, thereby giving the player the incentive to try to identify for each question what will be good data combinations and also to 'snatch' the best location for collecting each data type in order to maximize their points. There is an element of competition as once a player claims a type

of data for a particular square, the next player cannot claim the same data there.

This version of the game will soon be ready for testing. The evaluation will focus on two important aspects of the game. Firstly, to what extent do players learn and make better choices about placing data markers as the game progresses and they answer more questions. Do they try to claim more of the higher scoring squares for each data type? Do they match the data better to the question as the game progresses and they have the chance to answer more questions? Secondly, to what extent does the game help players to be more creative in thinking how to solve problems and design solutions from data? This will be assessed through pre and post-game creative tasks.

Adding Interactivity

The final stage, which is work-in-progress, is adding interactivity to the game. In this final version, a matrix of NFC tags will be embedded in the game board. This will allow a player to view data through a mobile device by holding the phone above the map on a square they are interested in, interfacing with the nfc tag located there. An app is being developed for Android-based smartphones. Figure 4 shows some of the interactive elements under development. (It should be noted that this photograph is for demonstration purposes. The NFC tags are lowered for the purposes of the photograph but they would be embedded in the game board in the final game). Calculations were performed to find how far apart tags had to be in order that they did not interfere with one another. A future iteration of the game may leverage other mobile sensor types such as visual signal processing. This may make for more flexible forms of interaction.

Figure 4. Testing the interactive game elements.

CONCLUSIONS AND FUTURE WORK

The interactive version of the game will connect directly to live data and will provide novel data visualisations to help users to 'see' data as it streams into their device. One data visualization that is under development will show *properties* of the data, including the volume and velocity of the data as it starts arriving after a sensor has been placed. This will be compared against a more traditional representation, such as time-series graph, to investigate which form provides the best data summary for making sense of the data. The aim is to investigate whether this strengthens the feeling that data is connected to an environment. A final iteration will investigate the possibility of moving beyond simple 'question-answer' interactions with data, and enable the user to download and analyse the datasets, thus increasing their expertise and data literacy even further. In the future, it is hoped

that it will be possible to develop a Datascape that links directly to real data hubs and can 'tie' any available geo-tagged data to a printed open source map. The future vision is to create an app that can recognize and geo-locate any printed map from, for example, streetmap which can be augmented by holding the phone above the map to show useful teasers and visualisations of data, thus revealing the datascape above a physical landscape through the device. Additionally, we plan to experiment with different levels of representation. By building 3 dimensional features into another iteration of the map and by giving it a topography we would increase the lifelike quality of the representation, bringing the symbolism of the map closer to the experience of reality. This would allow us to examine the relationship between different forms of tangible and virtual real-world simulation, and their impact on engagement with data analysis and design. In the process, we will be playing around with some of the broader philosophical and sociological ideas raised by Baudrillard with his notion of simulcra and simulcrum [1]. Initially though, our next step will be to complete both the non-interactive and interactive game versions and conduct user studies to evaluate the effect of interactivity on user engagement, the users' understanding of the data and their user experience more broadly. Informal observation of players during prototyping stages has so far suggested that the game will have the desired effect. We have seen that it can prompt lively discussion amongst players about different data types, how they relate to questions and how the different regions of the maps would be good or bad places to locate these types of data. We hope to be able to demonstrate that in the future, data can break free of the limitations of the database and be related more easily to the environment from which it comes. This will make the affordances of data for answering questions and for creating new data-drive products more salient.

REFERENCES

[1] J. Baudrillard. Simulacra and simulation. University of Michigan press, 1994.

[2] S. Follmer, D. Leithinger, A. Olwal, A. Hogge, and H. Ishii. 2013. inFORM: dynamic physical affordances and constraints through shape and object actuation. In UIST (Vol. 13, pp. 417-426).

[3] W. Gaver, A. Boucher, N. Jarvis, D. Cameron, M. Hauenstein, S. Pennington, J. Bowers, J. Pike, R. Beitra, and L. Ovalle. 2016. The Datacatcher: Batch Deployment and Documentation of 130 Location-Aware, Mobile Devices That Put Sociopolitically-Relevant Big Data in People's Hands: Polyphonic Interpretation at Scale. InProceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 1597-1607.

[4] C. Golsteijn, S. Gallacher, L. Koeman, L. Wall, S. Andberg, Y. Rogers, and L. Capra. 2015. VoxBox: A Tangible Machine that Gathers Opinions from the Public at Events. In Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15). ACM, New York, NY, USA, 201-208. DOI=http://dx.doi.org/10.1145/2677199.2680588

[5] C. Harrison, B. Eckman, R. Hamilton, et al. 2010. Foundations for Smarter Cities. IBM Journal of Research and Development 54, 4: 1–16. http://doi.org/10.1147/JRD.2010.2048257

[6] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian and K. Hornbæk. 2015. Opportunities and challenges for data physicalization. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3227-3236). ACM.

[7] M. Janssen, Y. Charalabidis, and A. Zuiderwijk. 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management 29, 4: 258–268. http://doi.org/10.1080/10580530.2012.716740

[8] The ODI. What is Open Data? Retrieved from https://theodi.org/what-is-open-data