Data Games

Marie Gustafsson Friberger
Malmö University
Östra Varvsgatan 11A
211 19 Malmö, Sweden
marie.friberger@mah.se

Julian Togelius, Andrew Borg Cardona, Michele Ermacora, Anders Mousten, Martin Møller Jensen, Virgil-Alexandu Tanase and Ulrik Brøndsted
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen, Denmark
julian@togelius.com, {aboc, merm, amou, mmje, viat, ugot}@itu.dk

ABSTRACT

We define *data games* as games where gameplay and/or game content is based on real-world data external to the game, and where gameplay supports the exploration of and learning from this data. This concept is discussed in relation to open data, procedural content generation and serious games, and research challenges are outlined. To illustrate the concept, we present six prototype games and content generators of our own making. We also present a tentative taxonomy of actual and potential data games, and situate the described games within this taxonomy.

1. INTRODUCTION

All games model the "real world" in some sense. For example, *Grand Theft Auto: Vice City* portrays the real and fictional Miami, flight simulators reproduce the experience and challenge of flying an actual plane with varying degrees of versimilitude, *Breakout* could be said to model ball games such as tennis in an abstract form, and *Chess* models military conflict in an even more abstract form. It is rare, but not unseen, for games to incorporate actual real-world data in such a form that the player could learn something about the world. Examples of the latter include the maps in some strategy games, and the vehicles in some driving and flying simulators. However, in most such games the real-world data has passed through a manual filter where human game designers select and reshape data in order to be suitable for the game in terms of e.g. challenge and overall aesthetics. It is very rare for a game to include a facility for the player to select their own real-world data from an external source. In this paper, we discuss the nascent idea of of games that allow the player to freely play with real-world data so as to explore and learn from it, and where all conversion of data from external sources into a form suitable for playing with happens automatically. We call such games *data games*[1] as introduced in [6].

The viability of this idea builds on two phenomena that have existed for some time but have only recently been organised into movements and research directions: *open data* and *procedural content generation* (PCG).

1.1 Open Data

Open data refers to data from governments, companies and other organisations that is made freely available on the Web in some form; recently a large number of local and national governments have made concerted efforts to make large amounts of geographical, demographical and economic data available to their citizens (and citizens of any other country) in order to increase transparency and citizen participation. Good examples of the latter are the government-run data portals such as that of the US[2] and UK[3] governments. Other organisations include the United Nations[4] and the World Bank[5], as well as cities, such as London[6]. Similarly, NGO's have made efforts to increase transparency and accountability through making detailed data publicly available about e.g. financial transactions in foreign aid.

But that a dataset is publicly available does not guarantee that it is easy to use, as it might be encoded in a proprietary file format or require extensive domain knowledge to understand. A variety of forms of access are used to provide open data to the public. One example is by providing the information in spreadsheets via government portals. This has drawbacks such as being cumbersome to parse and access. One way of overcoming this is by providing application programming interfaces (APIs).

Yet another way of providing access to open data is through linked data [1, 8]. Linked data is based on four principles: (1) using Uniform Resource Identifiers (URIs) to name things, (2) that HTTP URIs are used so that people can look up those things, (3) that the information returned upon lookup is useful and in a standard format – such as the W3C recommendations of Resource Description Framework (RDF) or the SPARQL Protocol and RDF Query Language

[1] http://data-games.org/
[2] data.gov
[3] data.gov.uk
[4] http://data.un.org/
[5] http://data.worldbank.org/
[6] http://data.london.gov.uk/

(SPARQL) – and (4) that links are provided to other URIs. In addition to some government data being available using these principles (such as parts of that of the UK[7]), there are many other initiatives, such as the DBPedia project [2], which makes the structured data found in Wikipedia's info boxes available as linked data.

The idea of linked data is closely related to the vision of the *Semantic Web*, which would allow for completely automatic knowledge extraction and discovery on the world wide web. However, in practice the vast amount of data now available to the public is not matched by similarly advanced and accessible tools to allow people to independently explore and visualise the data.

1.2 Procedural Content Generation

PCG refers to the automatic or semi-automatic creation of content for computer games through algorithms. "Content" here refers to levels, maps, items, quests, textures, rules, plants, puzzles and in general everything that is not the behaviour of in-game characters or the game engine itself. While the history of procedural content generation in games spans three decades (famous examples include *Rogue*, *Elite*, *Civilization* and *Diablo*), the last few years have seen increased use of PCG in both mainstream and indie games and the formation of an academic research community on the topic. Many different content generation problems have been proposed and studied, and methods that have been proposed include cellular automata [9], evolutionary computation [14], constraint solving [13, 12] and grammars [5]. In most cases, the content generation does not take any external factors into account (and is thus often and misleadingly referred to as "random" content generation) though there have been some attempts to base content generation on player behaviour or models thereof [7, 11]. The use of external data to feed procedural content generation is largely unexplored, except when counting examples of using actual maps as the basis for a partly procedural level generation in titles such as the *Microsoft Flight Simulator* or the racing game *Fuel*. Further, in these examples the external data typically acts as a backdrop to the game, with little constraining or enabling effect on player actions.

1.3 Serious Games and Procedural Rhetorics

The data games idea is also closely related to the ideas of serious, persuasive and educational games. *Serious games* are games which have other primary purposes than entertainment. The set of possible types of such games is understandably very large, and both the motivations for and shapes of serious games vary wildly. Two popular types of serious games are educational games, which intend to teach the player a particular curriculum (e.g. English grammar or how to mend a motorbike), and persuasive games, which try to persuade players to adopt a viewpoint or take an action, such as voting for a particular political party. In contrast, our concept of data games is empowering rather than normative: a data game is not meant to teach any particular skill or instil any particular belief, but rather to give the player the means to explore the data themself by tracing the consequences of their actions.

The theory of *procedural rhetorics*, as proposed by Bogost [3], is frequently used to understand and justify design choices in serious games, and can be used to understand data games as well. In procedural rhetorics, an argument is embodied in a game system, and the player is carrying out the argument by playing the game. For example, the game *September 12* is a statement about the futility of fighting terrorism through bombing supposed terrorists. The designers chose to make bombing terrorists a core game mechanic and affordance, and made the consequences of carrying out this central mechanic predictably disastrous, allowing the players to convince themselves of the message the game designer wanted to convey through interacting with the game system. In data games, the player could similarly convince themself of an opinion or fact through observing their own interaction with the game. The difference is that the game content is not designed by a game designer with a particular political perspective or learning goal in mind, but rather generated automatically based on freely available open data selected by the player. The goal is to allow players to explore the data and express and indoctrinate themselves as they see fit.

1.4 Data Games

Data games are games where gameplay and/or game content is based on real-world data external to the game, and where gameplay supports the exploration of and learning from this data. As yet, there are few games that are unequivocally sits within this definition. Some examples of games that are strongly related to our intended meaning are discussed below, in addition to games developed with the concept of data games in mind.

Research questions related to data games can be grouped into three themes: (1) Exploration, learning and playability, (2) Data selection and access, and (3) Game design and development. A fourth aspect, that of evaluation, is necessary to inform the previous three. Concrete examples questions to be explored include:

- How can we best explore data through games?
- Which themes, genres and game mechanics best support learning and exploration?
- Who wants to play a data game? What player types will be most likely to enjoy and learn from engaging with a data game?
- What data sources are useful and appropriate for data games?
- How can we access open data for game content?
- How can data be transformed into game content?
- How can we support the design of data games?

In this paper, we begin to address questions four through six by giving game examples and presenting a tentative taxonomy of data games.

1.5 Overview

We continue this paper by giving eight examples of data games. Then, a tentative taxonomy of data games is described and exemplified. Finally, we discuss open issues in the nascent field of data games.

[7] `http://data.gov.uk/linked-data`

2. EXAMPLES OF DATA GAMES

In this section, we will discuss examples of games that can be defined as data games according to the definition above. We focus on those created by ourselves. (The first two examples are implemented by the first two authors, the four following are student projects at the IT University of Copenhagen implemented by authors 3-8 under supervision of the first two authors.) While we have attempted to survey the literature, we make no claims to completeness. Below we also present two examples of games presented by others. In addition, games similar to our conception of data games can also be found in the work of Macklin et al. [10].

2.1 Open Data Monopoly

Open Data Monopoly is a board generator for the classic board game *Monopoly* that creates boards (and cards) based on demographic and geographic data [6]. Marie Gustafsson Friberger and Julian Togelius created Open Data Monopoly in order to explore the concept of data games through a proof of concept. In the current context, explaining this software in some detail can serve as a useful introduction to data games.

The board for the Monopoly board game consists of 22 streets, ordered by ascending "prosperity". In retail versions of Monopoly, the game board is modelled on an existing city, such as Atlantic City or Stockholm; the early streets on the board represent low-income neighbourhoods and the later streets represent glitzy locations with high property values. In Open Data Monopoly, each "street" on the generated board represents a city in the UK. The dataset that is being explored is a combination of geographical and demographical indicators from several sources, including National Indicators collected from local authorities by the UK government, geographical information from the ordnance survey and article lengths from Wikipedia. All data is freely available online and was collected and processed automatically.

Open Data Monopoly lets the player explore the dataset through defining their own concept of "prosperity". Arguably, different people have different conceptions of what makes a part of a city (or a city itself) prosperous, and the various factors contributing to prosperity do not always go together. High income, low (perceived or actual) criminality, high education, high or low ethnic diversity, high civic engagement...? The first step of Open Data Monopoly is for the player to select a set of such indicators and grade their contribution to prosperity from strongly positive to strongly negative. The game generator then generates a Monopoly board that reflects the player's definition of prosperity as well as possible, including extreme examples of high and low prosperity in the beginning and end of the game. (It also ensures that the locations that form the bases of streets are mostly well-known and evenly distributed around the UK.)

In this way, the generated game content acts as a lens through which the player can see demographics and geography of a country. Through expressing their own concept of prosperity, the player learns to understand both the underlying data and their own ideas of prosperity, and might even find reason to question and revise their ideas about what makes a city prosperous. In one experiment, the authors defined a multi-factor model of prosperity that resulted in a certain small resort town coming up top, where at least one of the authors would certainly not want to live. Another board generated with Open Data Monopoly is illustrated in Figure 1.

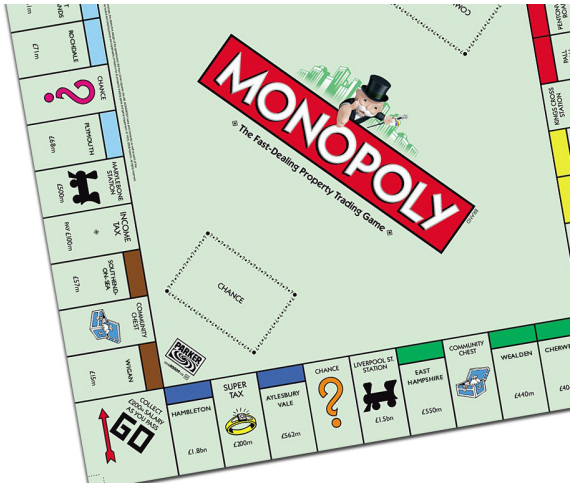

Figure 1: Open Data Monopoly.

2.2 Bar Chart Ball

Bar Chart Ball, by Julian Togelius and Marie Gustafsson Friberger, takes the ubiquitous bar chart and turns it into a ball game. Bar charts are commonly used to visualise numeric data, allowing the viewer to easily compare the magnitudes of several numerical variables. The basic design idea of Bar Chart Ball is to make data selection a core game mechanic, and not only part of the game content generation, in such a way that the player needs to learn about the data set in order to play the game better.

Our first implementation of Bar Chart Ball uses the same dataset as Open Data Monopoly, consisting of 72 demographic indicators for approximately 200 locations (cities and towns) in the UK. The playing field is a simple bar chart, where each of the ten bars represents a different location and the height of the bar represents that location's score on a selected indicator. For example, if the selected indicator is general mortality, City of London will have a short bar and a community with a high proportion of retirees will have a long bar, but for an indicator related to perception of drug users as a social problem the relation is likely to be the opposite. At any time, the player can select a different indicator, thus reshaping the "landscape" formed by the bar chart; the bars move swiftly, but not instantly, into their new positions.

Onto this landscape, a ball is dropped. The player's goal is to control this ball, and make it go where they want. However, the player's only tool for affecting the ball is changing the reshaping the bar chart through the judicious selection of indicators. If the player wants the ball to fall off the left end of the screen (a goal in the first version of the game), they might need to find an indicator that causes the bar immediately to the right of the ball to rise, pushing the ball left. Usually, getting the ball where you want requires a number of consecutive chart transformations. Playing Bar Chart Ball successfully requires being able to estimate the state of the chart after selecting a particular indicator, and thus the best way of getting better at the game is to learn to model the numerical relationships in the data underlying the game content.

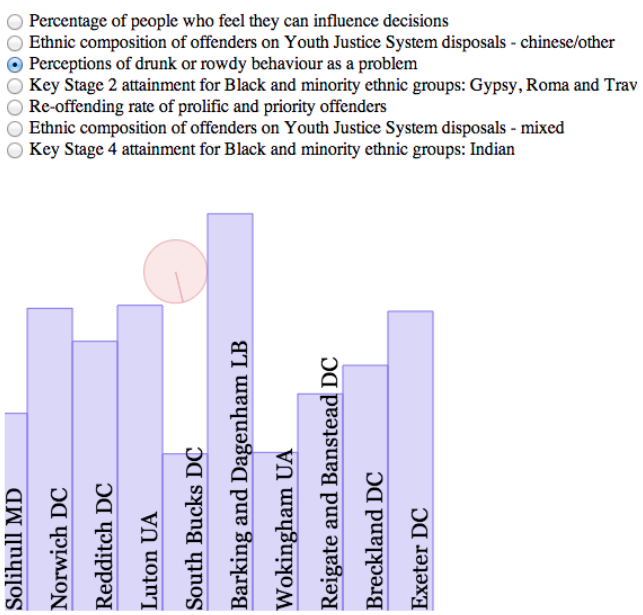

Figure 2: Bar Chart Ball. In this example, the player has selected a "Perception of drunk and rowdy behaviour as a problem" as indicator, and the ball is perched in the valley of South Buckinghamshire, next to the tower of Barking and Dagenham.

Bar Chart Ball is currently a prototype implemented in HTML5, using the Box2D physics engine. Figure 2 shows the current version of Bar Chart Ball in action. Ongoing development of Bar Chart Ball includes guaranteeing playability through optimisation of bar placement.

2.3 OpenTrumps

OpenTrumps is a content generator created by Andrew Borg Cardona for the popular card game *Trumps*. In Trumps, decks of cards focus on a particular type of entity (e.g. movie stars, sports cars, footballers) where each cards lists 4-5 attributes; if a deck is about sports cars, the features might be number of cylinders, top speed, acceleration and price. Players of Trumps draw cards and then choose on which feature to compare them, meaning that winning a game of trumps requires domain knowledge to be played well. OpenTrumps automatically creates balanced Trumps sets based on countries. It draws data from a UN database of countries and demographic indicators (life expectancy, education level, military expenditure etc), and uses evolutionary algorithms to create balanced decks of card. The user can choose which countries to include, and the software then chooses indicators that make sure each card (country) has a chance against most cards, or choose indicators and then the software finds countries that create a balanced deck. Figure 4 shows a card from such a deck.

2.4 OpenStreetRacer

OpenStreetRacer, by Michele Ermacora and Anders Mousten, lets the player select a geographical area from OpenStreetMaps, and the starting and ending point of a race track. It then generates two possible race tracks, one that takes the shortest route along existing streets from the start to the end of

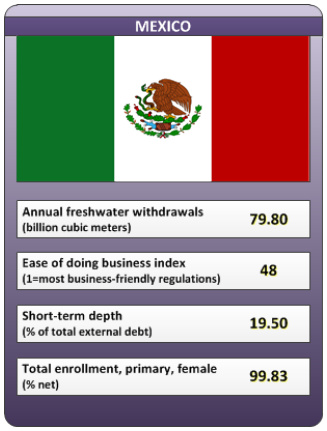

Figure 3: A card generated by OpenTrumps.

the race, and one that smooths the path to something more like a race track. The smoothed path makes smooth curves out of corners, and as a result the track cuts straight through buildings; the software simple reshapes the buildings to fit. The player can then select which track ("real" or smoothed) to play, and the software quickly generates simple 3D models of the buildings along the path, and lets the player drive a car from a first-person view. See Figure 4 for a screenshot.

2.5 Open Data Civilization

Open Data Civilization, by Virgil-Alexandru Tanase and Ulrik Brøndsted, is a map generator for the epic turn-based strategy game Civilization IV. The source data consists in high-resolution maps of the world from WikiMedia.org, where each map focuses on a different aspect (e.g. terrain, elevation, and the distribution of resources such as coal and oil). The player can then choose any area of the map, from the full map of the world to a map as small as the "greater" Copenhagen area. Based on the supplied map data, a playable Civilization IV is generated including terrain, climate and resources. An outstanding issue that is currently being addressed is how to handle the fact that the resource distribution of the real world is rarely conducive to balanced and fair gameplay. See Figure 5 for a screenshot.

2.6 Flight Leader

Flight Leader is a real-time strategy game by Martin Møller Jensen which tasks the player with guiding aircraft to their destination airports. The basic game content is the recorded flight paths of hundreds of aircraft, data which is gathered in real-time from the flightradar24.com website. However, that does not make for much of a game, as the vast majority of real-world flights end up at their destination airports. Therefore, a number of "ghost flights" are introduced, which might have no destination or be on a path to collide with some of the real flights. The player must then assign new destinations to real flights or to ghost flights, or put flight in holding patterns while waiting for destination airports to have landing slots free. The game is lost when the player has killed enough people by failing to spot and avoid impending collisions, and score is lost for landing a plane at the wrong airport. Unlike popular casual games with a flight controller theme, the gameplay in Flight Leader is based on real world data (including speed, altitude and passenger numbers of

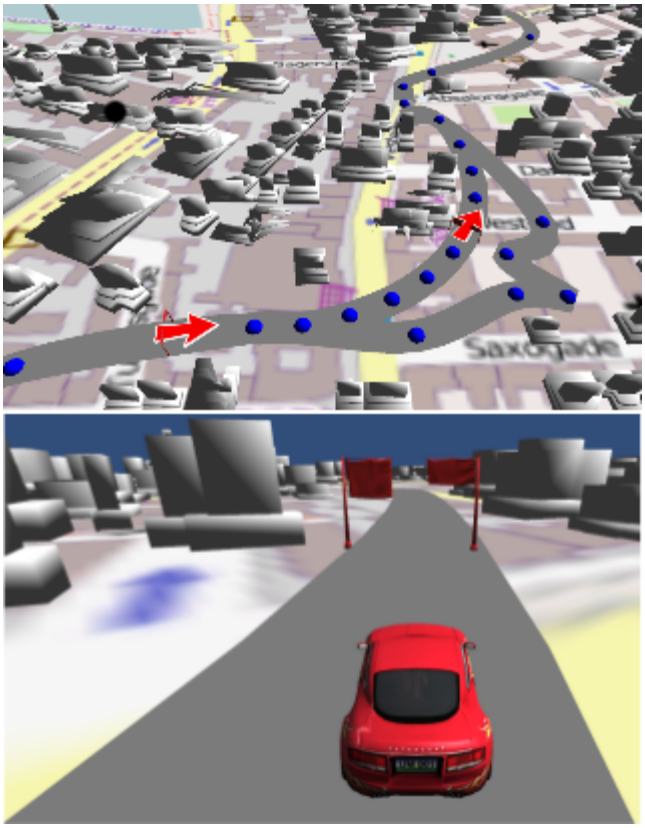

Figure 4: OpenStreetRacer. The upper screen shot shows two variations on a racing track through part of Copenhagen, and the lower screen shot shows the game's stock car driving on that track.

existing airplane types) and the player needs to understand the characteristics and constraints of actual flight traffic in order to play the game well. See Figure 6 for an illustration of Flight Leader in action.

2.7 Urbanopoly

Like our Open Data Monopoly, Urbanopoly [4] is based on the Monopoly game concept. It is more traditional adaptation, where the venues the player can buy are real places in the world. Available open data, such as OpenStreetMap, are used to generate the game storyboard. The main idea of Urbanopoly is to use the concept of games with a purpose [15] to verify, correct and collect geospatial data.

2.8 MuseumVILLE

Another possibility is to use cultural heritage content, such as images of paintings and artifacts. MuseumVILLE[8] aims to engage their users by letting them create their own museum based on his or her interests, using open cultural heritage content and data from Europeana[9]. The player takes the role of the museum curator, and composes exhibitions using material from Europeana. Points are scored by attracting visitors to the exhibition's page. In future versions, they propose that points also be scored from categorising material, adding a crowdsourcing dimension.

[8] `https://github.com/bogusjourney/museumville/#readme`

[9] Europeana is a portal for accessing digitised cultural heritage material, such as paintings and books, from more than 2,000 institutions across Europe.

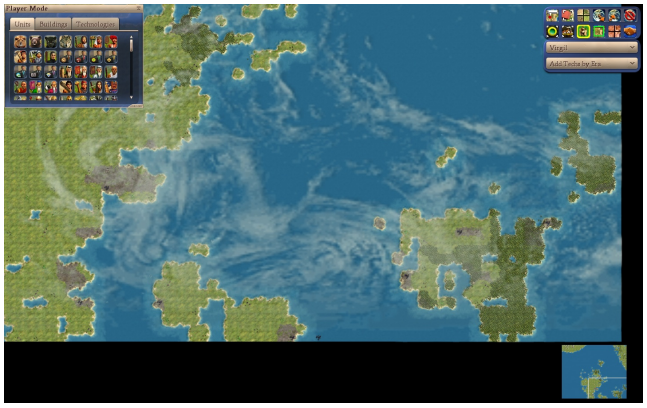

Figure 5: Open Data Civilization. The screenshot shows Fyn and parts of Zealand in Denmark as a map.

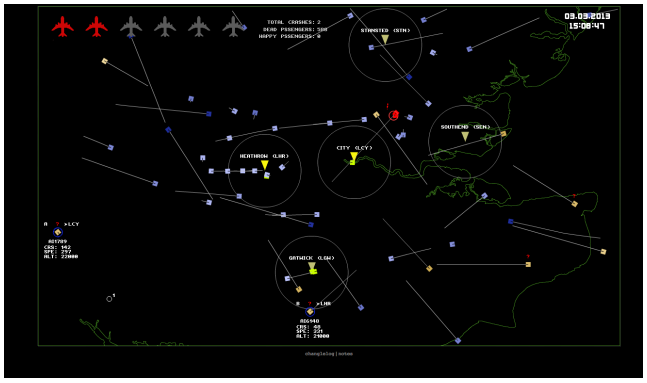

Figure 6: Flight Leader. The screen shots shows a number of planes over the Greater London area.

3. TENTATIVE DATA GAMES TAXONOMY

In [6] we suggest that possible data games can be categorised according to the following dimensions: underlying genre, source of data used in the game, and how the data is transformed into content in the game. To identify new areas of research and game development, this paper proposes a tentative taxonomy of data games, which expands and explicates the previously presented dimensions. As there is so little previous work, we have not been able to analyse a corpus of existing games to derive the taxonomy. Rather, we have taken our own game design ideas as starting points, and tried to envision combinations of game mechanics and data types that we know of. We expect the taxonomy to be revised and extended as the field of data games matures. We want to emphasise that this tentative taxonomy is meant only as a device for identifying promising research directions, not as a tool for classifying future data games.

Below we discuss and exemplify the dimensions, an overview of which is found in Table 1. For some of the values possible for the dimensions discussed in the previous section, we do not give any example values. This is because we do not know

of any data games with that characteristic, but we include them in the taxonomy because we find the values viable and can be explored in future games.

3.1 Game Genre

The underlying game genre forms a foundation for the data game design. If a familiar genre is chosen, the player benefits from familiarity with the game mechanics. Potential genres familiar to many players include board games, simple arcade-style computer games, first-person 3D computer games, trivia games and strategy games. For example, OpenStreetRacer is a first-person 3D game while Open Data Civilization is a strategy game. However, data games can also be created in genres that are not common or that combine familiar genres in new ways. Bar Chart Ball, for example, can be viewed as an example of a simple arcade-style computer game, but which may seem less familiar due to the innovative way of controlling the game.

For a lengthier discussion of how these genres can be used in data games and how they affect data needs and game play, see [6].

3.2 Data Source

Data sources are integral to data games and the data source shapes the nature of the data game. We have identified three subdimensions related to data source: data type, data topic and whether it is static or dynamic. Given the increasing feasibility of linking data sources, and the strength of procedural representation for visualising and allowing exploration of how different types of data interact and depend on each other, we expect many data games to draw on several different data sources of different types.

3.2.1 Data Type

The type of data determines what kinds of data games can be based on them. Types of data include numbers, text and images. For example, Open Data Monopoly and OpenTrumps both rely heavily on figures and associated text. In MuseumVILLE openly available art images from Europeana are selected by the user.

3.2.2 Data Topic

Some topics of data that are commonly made publicly available and are often in need of sense-making and visualisation due to their complexity and/or sheer size of the datasets are demography, politics, geography, energy, and infrastructure (such as transportation). Another topic, where complexity and size are issues in a different manner, is cultural data and content.

These topics have different affordances when it comes to understandability and game design. For example, many have an intuition for visualising geographic data, which can be used in many data games. However, there is a large amount of information and variable levels of description. Here, the games of Open Data Civilization and OpenStreetRacer give the player the option to reconceptualise known or unknown geographies in a familiar game.

Demographic data, on the other hand, is much harder to make sense of and the initial selection and combination of data can have great effect on interpretations. Open Data Monopoly gives the possibility to play with the value judgements in understanding demographic data. Bar Chart Ball and OpenTrumps offer a similar possibility of exploration, but the players' knowledge (or hunches) determines the game play ability.

Examples of using data topics of culture, politics and transportation can also be found in Table 1.

Though there are more topics available for data games than those described here, those that form the core of open data initiatives will probably be more prevalent in data games. Games may combine data from several different categories, and the category can tend to be more general or more specific depending on the game concept. One can also envision data games created for a general data source, where the player can generate a game for a specific specialised topic. Another possibility is to choose to combine openly available data sources with one's personalised data, for example, as part of a game in a social network. While this brings with it privacy issues, it also makes it possible to create games of high relevance to the player.

3.2.3 Static vs Dynamic

The rate of change in the data can both be an element of the game and affect the data selection. Data sources range from constantly updating, such as real time transport data, to those that rarely change, such as geographical and cultural heritage content. In between we have sources that are updated at regular but longer intervals, such as demographical and political.

For example, Open Data Civilization is based on geographical resources, which are more or less static. Open Data Monopoly is based on demographical and political data, which does change, but (usually) at a rather slow rate. Finally, Flight Leader is based on data which changes real time. Where along this scale of change the data sources are placed affects both the construction of the game and the type of game play experienced.

3.3 Data Selection

Both in the initial design of a data game and in the procedural generation of its content, data has to be selected. In many cases, the data source will already have been decided before a subset of that data is selected for use in an actual game session. Thus, by data selection we mean when and where this selection is made, who makes it, as well as how it is made.

3.3.1 Where and When?

One aspect of data selection is where and when the selection takes place. Two categories of data selection in data games are *game as visualisation*, where data is selected in the content generator, and *visualisation as game*, where data is selected inside the game. Almost all of the examples described above fall into the category of game as visualisation. Only Bar Chart Ball is a clear example of visualisation as game. Open Data Monopoly can be in both categories: the game board generation is visualisation as game, which the actual use of these boards for game play is game as visualisation.

3.3.2 Who?

In many data games, the player is active in data selection. The computer may also make some part of the selection (for example, when there is a randomisation element). One can also imagine a game setting where opponents select data. However, in the games described above, the player is an active data selector.

Table 1: The data games taxonomy exemplified. The list of examples for each dimension is not exhaustive with respect to the games described in this article.

Dimension	Sample values	Example
Game genre	board games	Open Data Monopoly
	strategy games	Open Data Civilization
	simple arcade-style	Bar Chart Ball
	trivia games	OpenTrumps
	first person 3D	OpenStreetRacer
Data source	datatype: numbers	Open Data Monopoly, OpenTrumps, etc.
	datatype: images	MuseumVILLE
	topic: geographical	Open Data Civilization, Urbanopoly
	topic: demographical	Open Data Monopoly, Bar Chart Ball, OpenTrumps
	topic: cultural	MuseumVILLE
	static	Open Data Monopoly, Open Data Civilization, etc.
	dynamic	Flight Leader (potentialy)
Data selection	where and when: game as visualisation	Open Data Civilization, OpenStreetRacer, etc.
	where and when: visualisation as game	Ball Chart Ball
	who: player	All examples
	how: geographical	Open Data Civilization, etc.
	how: value judgement	Open Data Monopoly
Data transformation	maps and boards	Open Data Monopoly, Open Data Civilization etc.
	levels	OpenStreetRacer
	rules	No examples
	cards	OpenTrumps, Open Data Monopoly
	dialogue	No examples
	enemies and NPCs	No examples

3.3.3 How?

How then does the player make a selection? One possibility is by selecting geography (Open Data Civilization and OpenStreetRacer). Another is by making some sort of value judgement, as in the indicator selection of Open Data Monopoly. The selection of works of art in MuseumVILLE can be motivated by both the desire an exhibition of interest to oneself and to score points with exhibition visitors.

3.4 Data Transformation

The final dimension relates to how the data is transformed into content in the game. Possible types of content include: Maps and boards, levels, rules, cards, dialogue, enemies and non-player characters (NPCs). The possible data transformations will depend on the chosen game genre, and different forms of data may lend itself to be transformed into different types of content. In the examples above, the most common type of transformation is into maps and boards, for example in Open Data Monopoly, OpenStreetRacer, and Open Data Civilization. Cards are generated by OpenTrumps and Open Data Monopoly.

It should not be assumed, however, that the best results are achieved by mapping the data to the type of content that best resembles it (such as geography to maps and economy to in-game economics). A powerful mechanism for sense-making may be be found in the change in perspective in inherent in transforming, for example, political arguments into spells.

4. OPEN ISSUES

In this paper, we have continued explicating the topic of data games. We now return to the three themes of research questions presented in Section 1.4.

4.1 Exploration and Learning in Data Games

The intention of data games is to promote exploration and learning. One part in examining this can be to analyse what topics of data are of most interest to explore, and which topics people usually have trouble getting an understanding of. A related aspect is to examine what genres and game mechanics best support different user groups and understanding of various topics of data.

As a first step, the games presented in this paper can be taken as starting points for experiments with various groups of users in order to evaluate how different data games are perceived when it comes to, for example, exploration, sense-making, learning, and ease of use (different aspects may be relevant for different games).

4.2 Data Selection and Access

Several challenges are posed by the underlying data used for game generation. One such challenge is that of data quality, which any application that builds on open data has to deal with.

Another challenge common to applications built on data is that of finding and cleaning the data. For example, in the Open Data Monopoly, we had to handle the fact that some indicators were not present in all geographic entities. One option to ease this aspect of the development of data games is to a common platform that can be used by several data games.

A related aspect is that of data "diversity". That is, finding data that is almost, but not quite, comparable. For example, where there is some small variation in the year for which it applies, what regions are used or what the under-

lying definition of an indicator is. This was the case for the UN indicators used for OpenTrumps. Here a decision has to be made about what is similar enough, and whether and how this should be communicated to the player.

Yet another issue related to data is that of access. Different methods can be used, from downloading an Excel file, to using an API to using SPARQL, each with their own challenges. For example, the data from the Excel file may not be the most recent, and the developer has to find out on his or her own when an update is made. When using SPARQL, while affording more complex queries, there may be performance issues. Depending on how the data is used in the game, local storage of the data retrieved via SPARQL may be a more stable option. The move to a web of more linked data will certainly make the development of data games easier.

4.3 Game Design and Development

The data games presented in this article are prototypes developed in a relatively unsystematic manner, and no user studies have been performed. It is very much an open issue how to develop a data game in a systematic manner. For example, would you begin with the data or the game mechanic? Once more playable data games have been developed, it needs to be further analysed how people play these games. For the design of data games, data collected during play can be examined to see what different patterns of play can be discerned. Also, preferred genres and transformations can be analysed.

There are several additional underlying issues to be explored, which will also be informed by the design, development and evaluation of more data games. For example, what real world data is playable, and through what transformation? What data is not only playable but also fun? What types of transformations lead to an unacceptable loss in veracity in relation to the original data source? In many cases, it will be necessary to use optimisation or constraint solving algorithms for data selection, and an open issue here is what kind of constraints or evaluation functions guarantee playability while being reasonably fast and keeping enough of the original data.

5. CONCLUSIONS

We have defined the concept of data games, which draws on open and linked data and procedural content generation to create games that allow players to explore and experience existing data in new ways. The possibilities of transforming data into game content are immense, and we are just starting to explore the ways in which such transformations can be meaningful. The six prototype games and content generators that this paper has given a brief overview of map only a very small part of the space of possible data games. We hope that these examples, along with the tentative taxonomy and identification of open issues, will get you started in thinking of your own way to incorporate open data in your game.

6. REFERENCES

[1] T. Berners-Lee. Linked-data design issues. W3C design issue document, 2009.

[2] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. Dbpedia - a crystallization point for the web of data. *Web Semantics*, 7:154–165, September 2009.

[3] I. Bogost. *Persuasive Games: the Expressive Power of Videogames*. MIT Press, Cambridge, MA, USA, 2007.

[4] I. Celino, D. Cerizza, S. Contessa, M. Corubolo, D. DellAglio, E. D. Valle, and S. Fumeo. Urbanopoly–a social and location-based game with a purpose to crowdsource your urban data. In *SoHuman 2012, 1st Workshop on Social Media for Human Computation, International Conference on Social Computing (SocialCom)*, pages 910–913. IEEE, 2012.

[5] J. Dormans. Adventures in level design. In *Proceedings of the workshop on procedural content generation*, 2010.

[6] M. G. Friberger and J. Togelius. Generating interesting monopoly boards from open data. In *IEEE Conference on Computational Intelligence and Games (CIG)*, 2012.

[7] E. J. Hastings, R. K. Guha, and K. O. Stanley. Automatic content generation in the galactic arms race video game. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(4):245–263, 2010.

[8] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 1st edition, 2011.

[9] L. Johnson, G. N. Yannakakis, and J. Togelius. Cellular Automata for Real-time Generation of Infinite Cave Levels. In *Proceedings of the ACM Foundations of Digital Games*. ACM Press, June 2010.

[10] C. Macklin, J. Wargaski, M. Edwards, and K. Y. Li. DATAPLAY: Mapping game mechanics to traditional data visualization. In *Proceedings of DiGRA*, 2009.

[11] N. Shaker, J. Togelius, and G. N. Yannakakis. Towards Automatic Personalized Content Generation for Platform Games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*. AAAI Press, October 2010.

[12] A. Smith and M. Mateas. Answer set programming for procedural content generation: A design space approach. *IEEE Transactions on Computational Intelligence and AI in Games*, 2011.

[13] G. Smith, J. Whitehead, and M. Mateas. Tanagra: Reactive planning and constraint solving for mixed-initiative level design. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(3):201–215, 2011.

[14] J. Togelius, G. N. Yannakakis, K. O. Stanley, and C. Browne. Search-based procedural content generation: a taxonomy and survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3:172–186, 2011.

[15] L. Von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.