Generating game content from open data

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

Julian Togelius
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen, Denmark
julian@togelius.com

1. DATA GAMES

We would like to add to the ever-growing list of things that can be done with games by proposing, discussing and exemplifying *data games*. A data game is a game that allows the player(s) to explore data that is derived from outside the game, by transforming the data into something that can be played with. In other words, games as a form of interactive data visualisation. We believe that games, through their procedural and interactive nature, can provide possibilities for visualisation over and beyond traditional methods. Data games differ from most other types of serious games in that they do not (intentionally) have an agenda; they are tools for the player to use to explore data with as few constraints as possible. This implies that the player, to the greatest extent possible, should be able to choose both what data to use and what in the data is important. Data games complement the open data and linked data movements, as the growing amount of data that is available in this way is a good fit for game-based exploration. As an example of a data game, we briefly describe *Open Data Monopoly*.

2. OPEN AND LINKED DATA

Open data, and especially open government data, is the idea of making data available to the public to enable, for example, transparency and reuse. Such data is provided by countries, cities and other international organisations. Domains for which data is available include finance, social and economic indicators, and community information.

A variety of forms of access are used to provide open data to the public. Spreadsheets are often used, but this has drawbacks such as being cumbersome to parse and access. One way of overcoming this is by providing application programming interfaces (APIs). Another way of providing access to open data is through linked data [1]. 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 (SPARQL) – and (4) that links are provided to other URIs. In addition to some government data being available using these principles, 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.

3. BASING GAMES ON DATA

There are a large number of ways in which games could be based on or incorporate open data in order to help make sense of, make accessible or provide a different perspective on data which comes from the "real world" outside of the game. The various types of possible data games could be categorised along the following dimensions:

1. Genre of game that would underly the data game. Possible game categories include board games, first-person 3D computer games, and strategy games, each with their own affordances and limitations.

2. Type of data used in the game. 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. Some categories of data which are often 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 geographic, demographic and political data.

3. How the data is transformed into content in the game. *Game content* refers to a large variety of types of artefacts used in games, including maps, levels, items, quests, rules, textures and music; almost any type of game content could plausibly be generated based on open data.

The type of contents available will of course depend on the game genre chosen, and different forms of data will map better or worse to different types of content. In [3], for example, the ancient flora of Manhattan is overlaid on the present-day cityscape. However, it should not be assumed that the best solution is to map the data to the type of content that best resembles that data (e.g. geography to maps and economy to in-game economics); the change of perspective inherent

in transforming e.g. political arguments into spells might be a powerful mechanism for sense-making.

Procedural content generation (PCG) refers to algorithmic methods for creating various forms of game content automatically or with only limited user input [4]. While different forms of PCG have been part of some published games since the early eighties, most of the early attempts suffered from very poor control of the output of the algorithms. Recently, efforts have been made to base content generation on models of player experience and to develop content generation methods that allow for human co-design, thereby allowing for better controllability of PCG algorithms.

Data games require the external data that is to be made explorable be converted into game content automatically, as the player is free to select what data to play. This means that we need to develop PCG algorithms that base their content on real-world data, or in other words data-based PCG algorithms. A large number of different types of such methods, based on different underlying algorithms, are imaginable; the choice will depend on both the type of data to explore and the type of content to be generated.

4. AN EXAMPLE: OPEN DATA MONOPOLY

Governments, such as that of the UK, publish national indicator data for local authorities. Examples of national indicators are *Self-reported measure of people's overall health and well-being* and *VAT registered businesses in the area showing growth.* For data published by the UK, there are around 200 indicators for around 350 unitary authorities, district councils, metropolitan districts, and London boroughs. Visualising this comprehensive data set is a challenge.

Monopoly is a board game where players buy streets and build houses on these. The aim of the game is to make money by owning streets and houses, and charging rent from other players who land on these streets; the winner is the player who still has funds when others are bankrupt. The game board consists of streets of increasing property values. In addition to streets, the game board has locations where the player gets to draw a Chance or Community Chest card, through which the player may e.g. gain or loose funds, or move to a different location on the board.

By using a common board game such as Monopoly as a visualisation metaphor, citizens can be provided with a means of visualising publicly available data about their country and neighbourhood. In this case, in addition to playing the game on a generated board, the generation of game content itself can provide insight into how different communities fare on different indicators. The main idea of the visualisation is to arrange the local authorities as Monopoly streets on the game board. The arrangement of the local authorities is determined by a weighting of different indicators and the values of each authority for these indicators.

Figure 1 shows the steps in generating a game board as well as chance and community cards. The first step is to preprocess the data from the UK open data initiative, mapping indicator values to geographical units and removing indicators with missing data. The next step involves an interface where the user selects what indicators to include and gives them a positive or negative weight. Based on this, the 22 streets of the game board and their costs are created. In the current version, values of streets are calculated in a simple fashion. First, an overall value between -1 and $+1$ is calculated for each geographical unit by summing the value of

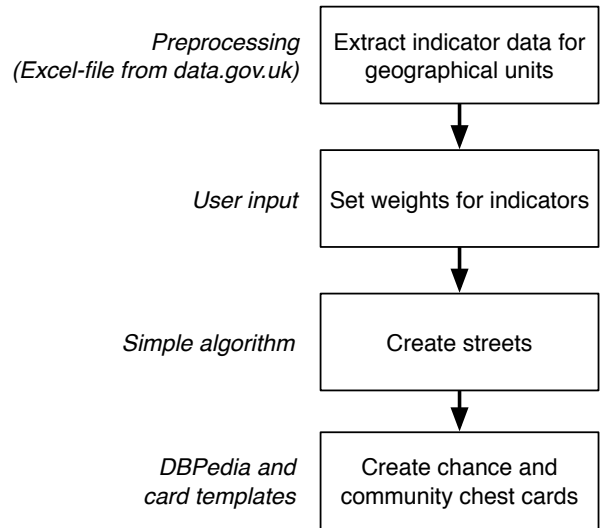| Preprocessing (Excel-file from data.gov.uk) | Extract indicator data for geographical units |
User input	Set weights for indicators
Simple algorithm	Create streets
DBPedia and card templates	Create chance and community chest cards

Figure 1: Flow of game board and card generation.

the selected indicators (multiplied by the weighting of the indicators) and dividing by the total number of selected indicators. The resulting value is then normalised to a number between 0 and 1000, which becomes the price of that street. As there are many more geographical units than available streets, we have to make a selection of which streets to include on the board. Currently this is done in the perhaps simplest way possible: the ratio of geographical units to streets is calculated as r, and every rth unit is selected, including the most and least expensive unit. The final step is to create cards based on the generated streets. In the prototype game, several templates for cards are created, which include variables to be filled when the game board is created, such as *You donate ?amount to the campaign of MP ?mp..* For creating the cards related to members of parliament (MPs), data from DBPedia is used. This is done through a SPARQL query using the name of the local authorities and other relevant properties.

5. FUTURE WORK

The next steps for the project include improving the street selection mechanism, extending the data used for card generation and evaluating the prototype.

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] C. Macklin, J. Wargaski, M. Edwards, and K. Y. Li. DATAPLAY: Mapping game mechanics to traditional data visualization. In *Proceedings of DiGRA*, 2009.
[4] 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.