

Exploration of Open Data through Procedural Content Generation

Gabriella A. B. Barros
Tandon School of Engineering
New York University
New York, USA
gabriella.barros@nyu.edu

Data games, procedural content generation, open data.

ABSTRACT

Procedural content generation for games has been studied extensively in the past (Shaker et al. 2015). In many ways, PCG often attempts to mimic a human's content creation process, and many creative methods have been proposed. However, little work has been done towards integrating open data into PCG. Games that attempt to do so have appeared recently, called data games. (Friberger et al. 2013). They refer to the use of data to automatically generate game content. Such content should represent the data in a way that both is engaging and understandable. This way, the user could interact with and learn from information that may be otherwise difficult or exhausting to engage with. Additionally, users identify easier with what they know: they may find more interesting to explore their hometown as an apocalyptic scenario, or to interact with abstractions of their idols in arcade games.

Therefore, this project aims at further exploring the space of data games. The main challenges of data games are data acquisition, transformation and balance. Data acquisition refers to how and where to obtain data. Transformation involves how to parse data into game content, and it differs depending on the data type, the content type and/or the game genre. Finally, balancing involves the task of delivering an enjoyable game experience, and ensuring that the original information is understandable, to some extent, while minimizing misinterpretation.

To achieve this project's goal, our approach consists of exploring how to generate different types of content, using different types of data. This involves the development of several projects of varying length. A general framework for data game generation has been proposed, as shown in Figure 1. It consists of two main parts: a **crawler** and a **parser**. The crawler is responsible for gathering and pre-processing data. The parser takes this data and transforms it into whatever game content is required. To generate different content, based on the same original data, it is only necessary to create a different parser. For example, the same information from Wikipedia can be used to generate items, dialogues or game levels.

Our first project attempts a somewhat direct transformation: creating map levels based on real maps. Generated levels could be played in *FreeCiv*¹, a game based on the Civilization series. OpenStreetMaps² provided geographical information that was parsed into the level

1. FreeCiv: <https://play.freeciv.org/>

2. <http://wiki.openstreetmap.org/>

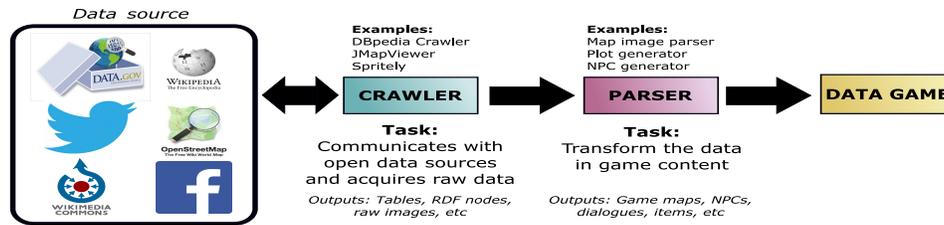


Figure 1: General framework for content generation in a data game.

topology, and images about natural resource deposits (e.g. oil, gold, etc) were manually obtained. This system applies image processing to place game resources on the level, and evolutionary algorithms to balance players' initial positions (Barros and Togelius 2015).

The second project, "Data Adventures", a work in progress, attempts to generate a complete digital adventure game using Wikipedia, OpenStreetMaps and Wikimedia Commons (Barros et al. 2015). In this game, the player is a detective and is given a crime to solve. The non-playable characters and locations are created from Wikipedia articles, and the relations between these articles forms the path between the victim and suspects.

There is still much to be done in order to achieve the goal of this work. The exploratory nature of this project requires a certain diversity of sub-projects. Firstly, we intend to improve the complete adventure game generation, especially gameplay-wise. A project of generating themed cards and decks for deckbuilding games is in design stage. Finally, in the future we also hope to integrate data game generation with game description languages, such as the Video Game Description Language. All these projects have different problems levels of difficulty. In the deckbuilding game, for example, we intent on reusing the crawler for Data Adventures, to see how it integrates with a different parser. But how can one map a Wikipedia article into attributes, such as strength and defense? On the other hand, for VGDL games, how to identify semantic meaning from the data, and map it into interesting mechanics? Alternatively, if we use a less "human-readable" source of data, such as government demographic information, can we parse it into more engaging representations, such as attributes in fighting games, or even unexpected content, like cloud shapes or clothes?

BIBLIOGRAPHY

- Barros, Gabriella A. B., Antonios Liapis, and Julian Togelius. 2015. "Data Adventures." In *Proceedings of the FDG Workshop on Procedural Content Generation*.
- Barros, Gabriella A. B., and Julian Togelius. 2015. "Balanced Civilization map generation based on Open Data." In *Proceedings of the 2015 IEEE Congress on Evolutionary Computation*. IEEE.
- Friberger, Marie G., Julian Togelius, Andrew Borg Cardona, Michele Ermacora, Anders Moustén, Martin Møller Jensen, Virgil-Alexandru Tanase, and Ulrik Brøndsted. 2013. "Data Games." In *Proceedings of the 8th Conference on Foundations of Digital Games*.
- Shaker, Noor, Julian Togelius, and Mark J. Nelson. 2015. *Procedural Content Generation in Games: A Textbook and an Overview of Current Research*. Springer.