

Domain Diagram and Analysis

Gabriel Gianordoli

Thesis Studio 2

Professors Sven Travis and Loretta Wolozin

March 10th, 2014

Concept Statement

Every second today, tech companies collect a large amount of records about people's online behaviours. Often referred to as "big data," these records are important in the study of society and culture. Some of this data is publicly available through the platforms themselves — pictures and comments from social media, for example. A considerable smaller part can be accessed through APIs. In both cases though, public access is limited and fragmented. As a consequence, its use is mostly limited to profit for the companies only.

This project uses Google aggregated search data as an exemplary case of this issue. It will build an archive of daily Google Autocomplete predictions, for multiple countries and products — web, Google Images, Youtube, etc. Users can have access to this information both through an online and a printed output. The former will provide an open tool for research, allowing multiple ways to explore this data. The latter will serve as a channel to communicate and discuss the findings of the project.

The purpose is to raise awareness on the cultural value of this data and answer questions such as: How can it be used in cultural studies? What information does it provide compared to other online sources? What are the the current limitations to its access?

Domain Diagram

The content-area subject of this project is *cultural big data*. It will approach the issue posed by the concept statement using *data visualization*, through both *online* and *printed media*.

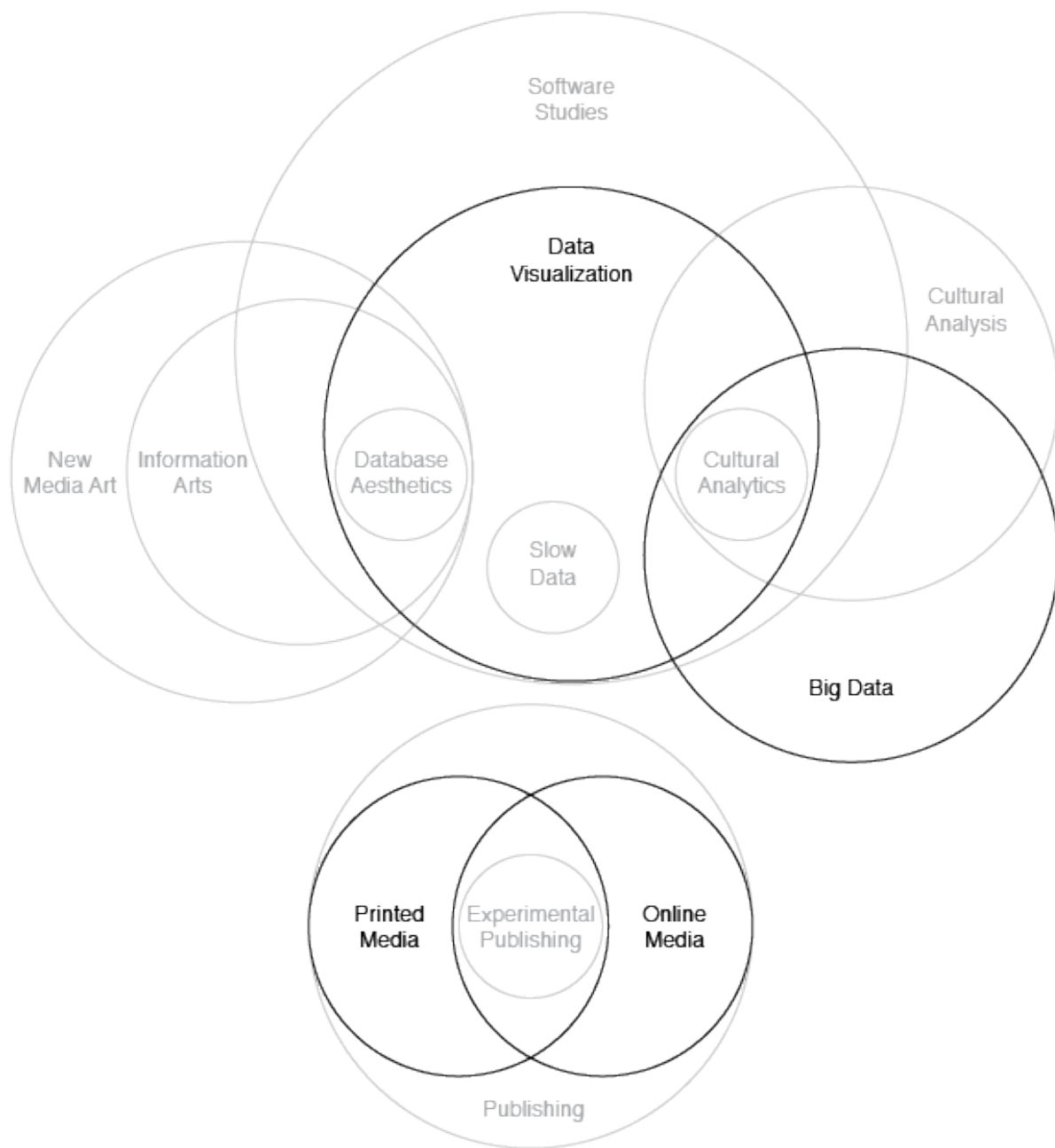


Figure 1: The core domains of this project (in black) and their relation to subsets and larger fields

Big Data

The term big data was first introduced in 2001, by analyst Doug Laney. [laney_3-d_2001] Technically, it is defined as “a broad term for data sets so large or complex that traditional data processing applications are inadequate.” (“Big Data” 2015)

Laney used the term in the specific context of business intelligence. Since then, the systematic collection of data has spread out to different areas. But the access to it has been limited to the companies. French philosopher Pierre Lévy states that “the big data available on the Internet is currently analysed, transformed and exploited by big governments, big scientific laboratories and big corporations.” (“Big Data Pierre Levy’s Blog” 2015) Lévy argues that social computing and decentralised communication led to the first digital revolution. According to him, a new revolution will happen when the processing and analysis of big data is made public.

Data Visualization

This project defines *data visualization* as the process of giving visual form to data. The concept should not be mistaken by the final visual piece, which is only the last part of a sequence. As data artist Jer Thorp points out “people talk about visualization as a thing(...) A visualization is a noun but it’s a verbal noun, it has a verb built into it. Visualization is a process.” (“NG Live! Jer Thorp: Numbers That Paint the Picture” 2015)

This idea is also present in the framework described by data visualization expert Ben Fry, which lists seven stages for visualizing data: acquire, parse, filter, mine, represent, refine, and interact. [fry_visualizing_2008]

A broad range of practices can fall into the *data visualization* category though. This project borrows methods from at least 3 of them: database aesthetics, slow data, and cultural analytics.

Database Aesthetics *Database aesthetics* here is considered a subset of information arts, on the intersection with data visualization. It defines works of art that apply the logics and aesthetics of databases, often providing “a way of revealing (visual) patterns of knowledge, beliefs, and social behavior.” (Paul 2007)

Information arts is defined by scholars Christiane Paul and Jack Toolin as works that “explore or critically address the issues surrounding the processing and distribution of information by making information itself the medium and carrier of this exploration.” (Paul and Toolin 2015)

This project will apply principles from this practice in its printed form. The idea is to make visible the critical aspects of the research.

Slow Data The term *slow data* was first used by business intelligence analyst Stephen Few, as a reference to *slow food*. While the culinary movement was a response to the spread of fast-food, Few's target is big data: "I believe that it is time to extend the Slow Movement to the realm of information technology. In this time of so-called big data, too much is being missed in our rush to expand." ("Visual Business Intelligence – The Slow Data Movement: My Hope for 2013" 2015)

Few argues for a better decision-making process based on big data. His principles apply directly to business intelligence. In this project, slow data is a deliberate strategy to engage users in a critical exploration of data. More than highlight findings, it will serve as a channel for reflection.

Cultural Analytics The expression *cultural analytics* was introduced by researcher Lev Manovich to define both the subject and techniques applied by him and his collaborators in the *Software Studies Initiative*. It argues for a "systematic use of large-scale computational analysis and interactive visualisation of cultural patterns," (Manovich 2009) utilising of large digital datasets.

This project shares some of the principles of *cultural analytics*, approaching big data as an important cultural record. However, the *Software Studies Initiative* uses mostly media publicly available — digitised works of art or pictures from social media. This project proposes the addition of aggregated search data as an asset in cultural research based on big data.

Writing and Designing Iteration

The definition of the domains helped reframe the concept statement of this project. The previous one had a focus on the technical aspects of the subject — databases. However, the problem with not knowing how they work is just part of a larger issue, that I found out researching the subject area. What do they refer to? Why are they important? Why aren't they accessible?

As a consequence of these new questions, the online and printed outputs now have better defined roles and audiences. I will develop the former as a tool, with focus on research and exploration. The latter will be a communication piece, serving as a report of my process and a way to discuss the importance and limits of using big data in cultural studies.

Bibliography

“Big Data.” 2015. *Wikipedia, the Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Big_data&oldid=650299928.

“Big Data Pierre Levy’s Blog.” 2015. Accessed March 8. <http://pierrelevyblog.com/tag/big-data/>.

Manovich, Lev. 2009. “Cultural Analytics: Visualising Cultural Patterns in the Era of ‘More Media’” *Domus, (Milan), March*. http://manovich.net/content/04-projects/061-cultural-analytics-visualizing-cultural-patterns/60_article_2009.pdf.

“NG Live!: Jer Thorp: Numbers That Paint the Picture.” 2015. Accessed March 8. <http://video.nationalgeographic.com/video/ng-live/thorp-data-lecture-nglive>.

Paul, Christiane. 2007. *The Database as System and Cultural Form: Anatomies of Cultural Narratives*. na. <http://visualizinginfo.pbworks.com/f/db-system-culturalform.pdf>.

Paul, Christiane, and Jack Toolin. 2015. “Encyclopedia of Library and Information Sciences.” Accessed March 8. <http://www.tandfonline.com/doi/abs/10.1081/E-ELIS3-120043697>.

“Visual Business Intelligence – The Slow Data Movement: My Hope for 2013.” 2015. Accessed March 7. <http://www.perceptualedge.com/blog/?p=1460>.