

Big Data and Their Impact on Libraries

Rafael Ball

ETH Library, ETH Zurich - Federal Institute of Technology Zurich, Zurich, Switzerland

Email address:

rafael.ball@library.ethz.ch

To cite this article:

Rafael Ball. Big Data and Their Impact on Libraries. *American Journal of Information Science and Technology*. Vol. 3, No. 1, 2019, pp. 1-9. doi: 10.11648/j.ajist.20190301.11

Received: November 29, 2018; **Accepted:** January 14, 2019; **Published:** February 22, 2019

Abstract: Academic libraries have a large amount of primary and secondary data with academic content. These contents can be linked and supplemented with freely available data on the Internet and generate benefit for their customers. This includes, for example, the date for indexing of library holdings as well as customer data and their media usage for the development of new helpful services. Thus, methods and knowledge of Big Data applications are required. With the help of Big Data technologies, these added values can be created. However, this presupposes that the limitations and possibilities of Big Data technology are being taken into account and that correlations are accepted as sufficiently accurate. Especially classical librarians have to cut back on their accuracy requirements. This paper gives an overview of the possibilities and chances of using large data amounts in libraries, presents hypotheses and explains practical examples.

Keywords: Big Data in Libraries, Classification, New Library Services, Data Management

1. Introduction

The digitisation of our world is advancing relentlessly and a far-reaching and serious side effect of this transformation is the fact that humankind is producing ever more data. Thanks to technology such as faster broadband connections, mobile

technology, ever more powerful computers, larger memories and the Internet of Things, the data volume is set to increase exponentially in years to come. For instance, the market research company IDC envisages a data volume of over 160 zettabytes by the year 2025; it currently amounts to approximately 25 zettabytes.

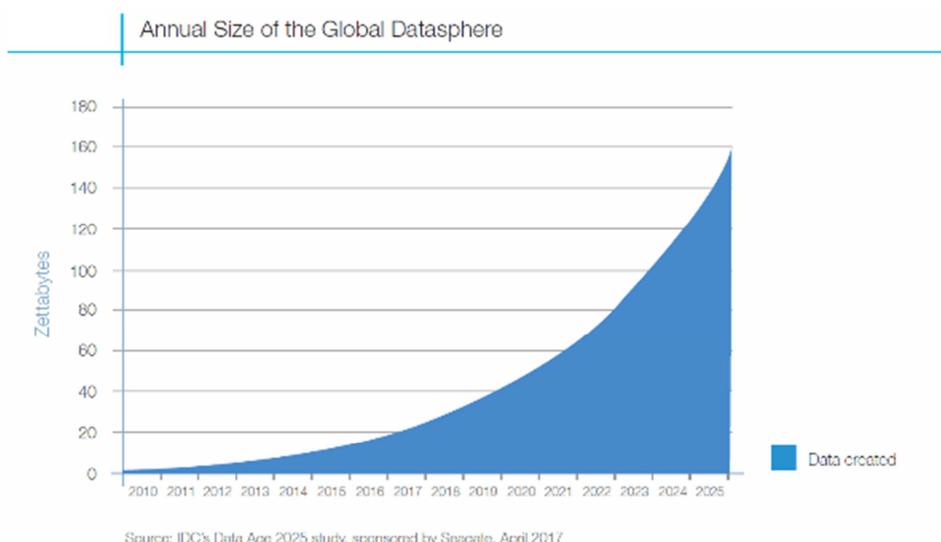


Figure 1. The development of the global data volume until 2010 (source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017 (available online: <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf> site visited on 06.11.2018), Page).

Merely saving data does not do justice to the possibilities that this form of information has to offer. In this context, many like to talk about the “oil of the 21st century” through the possibilities of data analysis. As Doug Cutting, who developed Hadoop, the open-source software framework by Apache, believes: “*Big data and open source will influence virtually every industry as big data helps us understand the customers better.*” Meanwhile, this is regarded in a similar way in the library and information system, even if it is not yet really implemented. Librarians, information scientists and other information professionals are becoming more interested in the subject. The growing significance of big data is evident in the diverse library conference programmes in the last five years. Talks or entire focus areas on big data can always be heard here.

In principle, libraries have not closed their minds to electronic digital media; on the contrary, they embraced them in their portfolios extremely early on and interpreted them as part of their collections and services. In the key indexing area of media and contents, however, they have clung to the accuracy paradigm of the cataloguing mindset of the 19th century to this day. This renders it virtually impossible to even contemplate the new technology and methods of big data that have been used successfully in many sectors for a number of years now. Here, big data technology is especially of the utmost importance in connection with indexing and providing library media, contents and data.

John Bailey from Dell Company drove this home in his quote:

Institutions that fail to embrace big data as an opportunity will to be left behind by those that do as the power of data-informed strategic decisions fuels them forward in the league tables. [1]

As far as the use of digital data is concerned, big data currently has the strongest implications in its impact on the information and knowledge society. While the details are not clear as to the where the big data topic will take us, we can agree that its use will have a profound effect, not just when it comes to handling digital information and communication, but also for the correctness of statements and their generation.

After all, big data describes a phenomenon that generates statements based on a vast quantity of data that do not signify any causality. Pure correlations emerge that permit statements, and the truth of their substance increases with the amount of underlying and evaluated data.

As only “small data” has been available until now (because measuring and recording it was time-consuming), libraries in particular wanted to evaluate the little data as precisely as possible.

For instance, books and journals have always been catalogued with the utmost precision. The focus was on recording all possible and achievable formal and content-related parameters and data accurately. These were slotted into a precise grid in accordance with extremely detailed rules and record-keeping regulations and accepted.

In doing so, libraries always used (well) structured metadata. The classic catalogue data as a reference for the holdings was the core element upon which the cataloguing and usage services were based. The more precisely it was collected and recorded, the better. Consequently, there were entire schools of cataloguing philosophies and ideologies, and a veritable conflict of opinions regarding the best description model. These altercations regarding the right cataloguing systems and rules still exist to this day.

With the advent of electronic data processing, the classic cataloguing model was transferred to the computer world: the relational databases were just the ticket for this goal. As a result, the analogue rules were transposed to the world of computers one to one and created extensive classification systems: after all, only what had been entered in the correct field in the relational database could be found in the search. From a database perspective, this was also absolutely correct at the time. And therefore, a person could only really search correctly in a database if they had filled it themselves. The more precise the entry in the database now was, the more complicatedly the fields were described and the more extensive the metadata set was, the more accurate the search thus had to be. The accuracy ideology of librarians was transferred from the analogue world to the early computer world.

Since the mid-1990s not only has electronic catalogue data been available, but also digital contents in a vast range of forms. Besides highly structured meta and catalogue data, which libraries have only had dealings and experience with thus far, increasingly vast quantities of unstructured data therefore also exists. This data is increasingly being supplemented with information that is freely available on the web and not only triggers changes in user behaviour, but also harbours tremendous potential for libraries.

However, the search for information and literature, especially in the STM subjects (science, technology and medicine), is no longer conducted according to the systematics or search logic of libraries, but rather online “Google style”.

The library inventory culture and the users’ search behaviour have evidently developed in different directions.

Nowadays, libraries go to great lengths to record a large number of extremely different categories in the reference system, systematise them and (still to this day) squeeze them into the logic of a relational database. The mental logic of a library user of the 21st century from Generation Y and Z is ousted just as much as the technical possibilities offered by big data and its algorithms today.

These days, academic libraries are confronted with an enormous amount of structured and unstructured data. Library work no longer focuses on books, journals and catalogue data, but rather all manner of unstructured and structured data and data forms: texts, metadata, images, audio files, videos, research data, 3D digital copies and software. Referencing, indexing and displaying these contents, however, can only succeed in keeping with the times now if

state-of-the-art indexing techniques from the field of big data are used instead of library methods that follow the logic of the 19th century. The notion of library accuracy and a tragic argument by analogy on the duties from the past still lead to vast and detailed metadata being generated and the digital objects – like analogue books – being indexed.

1.1. What Exactly Big Data Is

Big data is a relatively new concept that has caused quite a stir in the media in recent years. In extremely simple terms, this is taken to mean the storage, management and evaluation of vast quantities of data, which might be available in a wide variety of forms, i.e. as structured and/or unstructured data, and in diverse file formats. In principle, this is described with the three components:

- volume- variety and- velocity at which the data is generated, evaluated and processed.

Meanwhile, some authors add veracity/validity (referring to dubious data and ensuring the quality of the data) and value (referring to the generation of corporate value from big data) to these three Vs.

Wikipedia defines big data as follows, for instance: “Big Data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them.” [2]

In principle, however, the topic of big data is still something of a “black box”. The marketing research company Gartner, for instance, has observed a major interest in the topic. Seventy-three per cent of the companies polled in this study have already invested in big data solutions or are planning to do so in the foreseeable future. By their own account, however, only 15 % have managed to use and realise big data projects thus far. Difficulties such as finding sufficiently qualified staff, legal uncertainties, risks due to data protection issues, such big data projects taking too long to realise and the development of new big data solutions being too quick, i.e. existing ones rapidly becoming outdated, are mentioned as the biggest challenges for using big data productively. [3]

Another problem with big data is the vagueness regarding the amount that actually constitutes big data. As there is no fixed or standardised answer to this question, the marketing research company IDC, for example, conducted a survey among 254 German companies in September 2012. The study *Big Business dank Big Data? Neue Wege des Datenhandlings und der Datenanalyse in Deutschland 2012* (“Big Business Thanks to Big Data. New Ways to Handle and Analyse Data in Germany 2012”) clearly reveals sometimes heavily divergent perceptions regarding what constitutes big data. Thirty per cent are of the opinion that we can speak of big data when the data volume lies between 100 and 1,000 terabytes (TB). A similar proportion regard this as being between 50 and 100 TB, 17% from 10 to 50 TB, and around 15% only upwards of 1 petabyte (1,000 TB). However, there are also companies that class lower data volumes (below 10 TB) as big data.

1.2. Big Data in Libraries

There is a host of reasons as to why libraries should examine the topic of big data more closely. These not only include the departure from the notion of accuracy in cataloguing, but also the opportunities in processing academic data and the link to free data on the web.

1. Nowadays, vast quantities of data are created in virtually all disciplines, e.g. by physics experiments at CERN [4], satellites in the Earth’s orbit, genetic engineering, the health sector or market research surveys. However, we also find big data in libraries, museums or archives. More and more researchers are looking to use the collections there as a whole to analyse data and organise information in a completely new way. Therefore, many libraries have already been in the big data business for some time – unfortunately, often unbeknown to them. Large libraries particularly have an almost unmanageable amount of data owing to their digitised collections. 2. Big data affects libraries directly as they could use big data tools to analyse their large data holdings, such as to understand their own users better and thus be able to offer new or improved services. 3. Big data affects libraries indirectly as academics at universities will rely on big data increasingly more frequently in their research. [5] 4. For economic reasons as big data can also lead to cost reductions, automations and faster and better decisions, for instance. 5. Examining big data is necessary and harbours opportunities. Currently, only 0.5% of all the data worldwide has been analysed, i.e. there are still plenty of opportunities to get involved in big data projects. [6]

All of these reasons should prompt libraries to tackle the topic of big data precisely in practice, especially since pure information science seems to have moved away from this subject as... *the institutes for information science, big data, digital humanities etc. (are springing up) everywhere. In this respect, information science is expanding massively – only present information scientists are struggling to hold their own in this environment.* [7] In the current NMC Horizon Report Library Edition 2017, big data is incidentally referred to as one of the six key technological developments for libraries, albeit stipulating a short-term implementation time, namely in the space of a year. Nonetheless, the demand among users for such new big-data-based services is also still rather thin on the ground. In Marco Humbel’s Bachelor’s thesis *Die Umsetzung von Open Data an Wissenschaftlichen Bibliotheken in der Schweiz* (“The Implementation of Open Data at Academic Libraries in Switzerland”, [8] expert Michael Ehrismann, for instance, is asked whether, “*The acquisition of metadata and data is already available to externals?*” His response on e-rara.ch, the platform for digitised books from Swiss libraries: “Nobody has asked regarding big data analyses yet.” [9]

Therefore, the use of big data seems to be a relatively new field in libraries at the moment.

1.3. Fields of Application for Big Data in Libraries

In principle, three main fields of application can be

distinguished for big data: data as sources, data analyses (i.e. collecting, cleaning up, integrating and processing), data visualisation (presentation and communication).

The starting point for the following considerations is the aforementioned fact that libraries possess a vast quantity of data in the form of books, journal papers and studies in both physical and electronic forms. The library holdings were originally intended for researchers or public users to find and access individual pieces of information they require. Thanks to the digitisation of the collections, however, it is now possible to study this library data as a whole or in parts for data mining for using other data analysis methods.

Library analytics manager Aaron Tay, who works at the University of Singapore, therefore cites five reasons as to why data analyses in libraries (library analytics) will spread in the near future: - increasing general interest in big data, data science and artificial intelligence. - library systems are becoming increasingly more open and efficient as regards the use of analysis methods. - nowadays, academic libraries are forced to demonstrate their value for their host organisation or the general public. Data analyses have the potential to achieve precisely this. - mounting interest in learning analytics, i.e. the possibility to evaluate data from learners/students with a view to improving the learning environments and processes. - the increasing academic focus of research data makes synergies possible. For academic libraries, the focus will no longer be on open access in future, but far more on open data, if not open science. [10]

2. Examples

2.1. Library of Congress: Creating and Providing a Twitter Archive

One of the most famous, most spectacular and earliest examples of the use of big data in a library environment is undoubtedly the partnership between the short message service Twitter and the largest library in the world, the Library of Congress (LoC), announced in 2010. The goal was supposed to be to archive and retain every tweet ever tweeted. Unfortunately, this project proved trickier than initially anticipated. When the project was launched in 2010, approximately 55 million tweets were tweeted daily via Twitter. Meanwhile, the figure has ballooned to over 500 million tweets per day. Ever since, however, researchers have been waiting in vain to be able to evaluate this wealth of data. It is currently unclear whether this will ever be the case as the LoC has been unable to render this Twitter archive publicly accessible thus far. Representatives of the Library of Congress are also currently unable to make an estimation as to whether and when this might ever actually be the case. Irrespective of this, it is an impressive example of the potential of big data for libraries, especially if we consider that an institution which is over 200 years old is able to cooperate with a start-up that was only four years old at the time thanks to big data. [11]

2.2. Creation of a Metadatabase for Geophysical Data in Australia

The Business Information Survey from 2013 [12] reveals that information professionals have more or less been left out in the cold as regards the major fashionable topic of big data. Only a few of the respondents polled are evidently involved in such projects. However, they were involved in a specific example from Australia. The goal was to establish a metadatabase using geophysical data from the oil concern Shell in Australia. The realisation of this big data project also required the expertise of Australian librarians. Geophysical data is very difficult to manage as it involves complex “big” data that cannot be recorded using traditional cataloguing. Specifically, this concerns petabytes of data in a wide variety of file formats, media forms and licensing conditions. This ranges from raw data to processed and interpretable data. During the project, the librarians collaborated with geophysicists, geophysical data analysts, IT experts and database developers. The information specialists were responsible for the following tasks: - developing the necessary metadata fields including consultation- developing a controlled vocabulary and name conventions- defining the necessary search parameters- highlighting possibilities for additional functionalities- testing the database including feedback- importing metadata- developing user guidelines-offering training courses This case study reveals that big data projects do not have to be restricted to their own holdings for information specialists. Instead, librarians especially have key expertise for big data applications in the field of metadata [13, 14, 15].

2.3. Big Data Applications for Books (Harvard University Library)

In April 2012 Harvard University Library, the largest university library system in the world, began publishing all its metadata on over 12 million materials, such as books, videos, audio recordings, images, manuscripts, maps and other content. Due to copyright regulations, naturally it is not possible for the library to render all these materials freely accessible in full text online. Nevertheless, the metadata already constitutes a valuable treasure trove of data, which the Co-Director of Harvard Library Lab, David Weinberger, describes as “big data for books”.

The University of Michigan had already conducted a similar project in November 2010 [16]. Based on the quantity of data, however, these two cases cannot really be referred to as big data. For instance, the amount of data published by Harvard Library totals at around 4 GB [17].

As a result, there is fresh criticism of the use of the term “big data” in library contexts when many data analysis projects in libraries bandy around the expression even though the quantity of data is negligible. [18, 19].

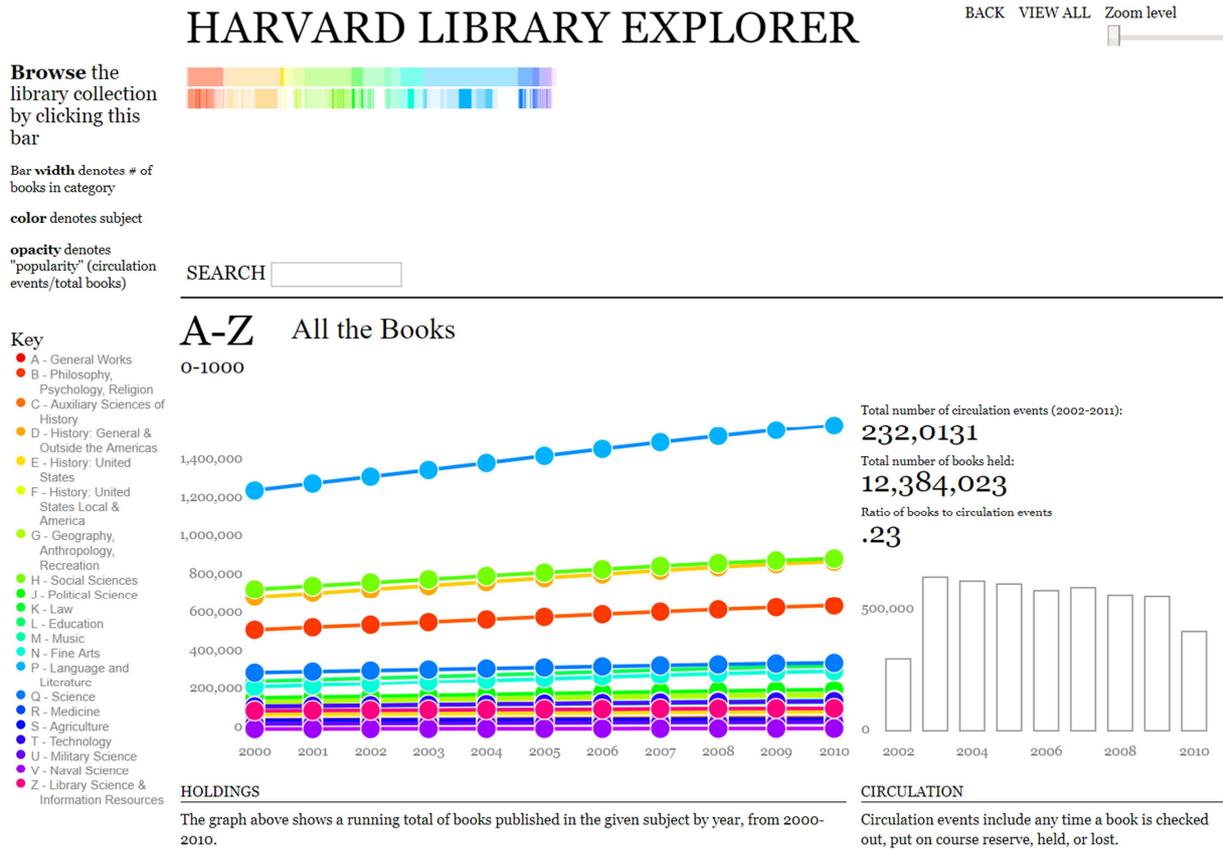


Figure 2. Harvard Library Explorer. (source available online: <https://lil.harvard.edu/toolkit/> site visited on 06.11.2018).

2.4. Jisc & HESA Library Data Labs Project

The non-profit educational institution JISC (Joint Information System Committee) [20] and HESA (the British Higher Education Statistics Agency) joined forces to provide institutions of higher education and other non-profit organisations in Great Britain with an efficient platform with various tools so they can analyse large quantities of data. Officially launched on 30 November 2015, this analysis platform is called Heidi Plus. [21] The idea behind this web-based platform is to hand decision-makers a tool to save time and money on the one hand and gain swift and easy access to the information they need.

JISC and HESA have now launched a special project called the Library Data Labs Project that also enables libraries to analyse large quantities of data using the Heidi Plus platform. This project differs from previous uses of Heidi Plus: its goal is to answer library-related questions. Over a three-month period, five inter-institutional teams comprising 23 different university libraries and a JISC team began answering certain questions from their libraries through data visualisations. The data sources evaluated for this project include SCONUL, Ulrich, Dewey, Altmetrics, H-Index, IMD etc.

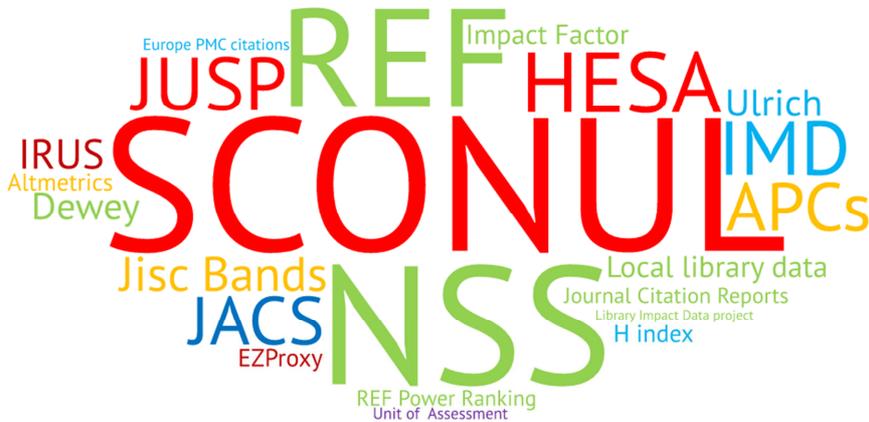


Figure 3. Data sources for the HESA project. (source available online: <https://libraryservices.jiscinvolve.org/wp/2016/10/library-data-labs-project/> site visited on 06.11.2018).

Essentially, it is a business intelligence project based on big data.

The issues examined within the scope of this project included the following: - the impact of the library premises on the students’ satisfaction and the various ways in which the buildings are used by various groups. - comparison of the library editions for a particular topic to find out what an “outstanding performance” looks like in this field. - analysis of the value of electronic journal subscriptions by considering the usage and costs, and involving the impact in the possible exchange of titles. - the initial inclusion of usage data from all participating institutions and the combination with other data sources provide a considerably more accurate image of the individual usage frequency of journals. [22]

2.5. Brooklyn Public Library (BPL): Big Data for the Visualisation of User Data

The big data project conducted by Brooklyn Public Library (BPL) pursues two primary goals: being able to make swifter and more data-based decisions. As this has not been the possible until now due to the existing dependence on external advisers and outdated reporting systems, the BPL turned to Tableau [23], an interactive data visualisation solutions provider. The BPL is a major public library in New York with numerous sites and branches. Consequently, data is recorded in many different places and systems. As a result, it was difficult to find out where the data required comes from. Once this Sisyphean task had been completed, the “magic” of Tableau’s data visualisation could take effect.

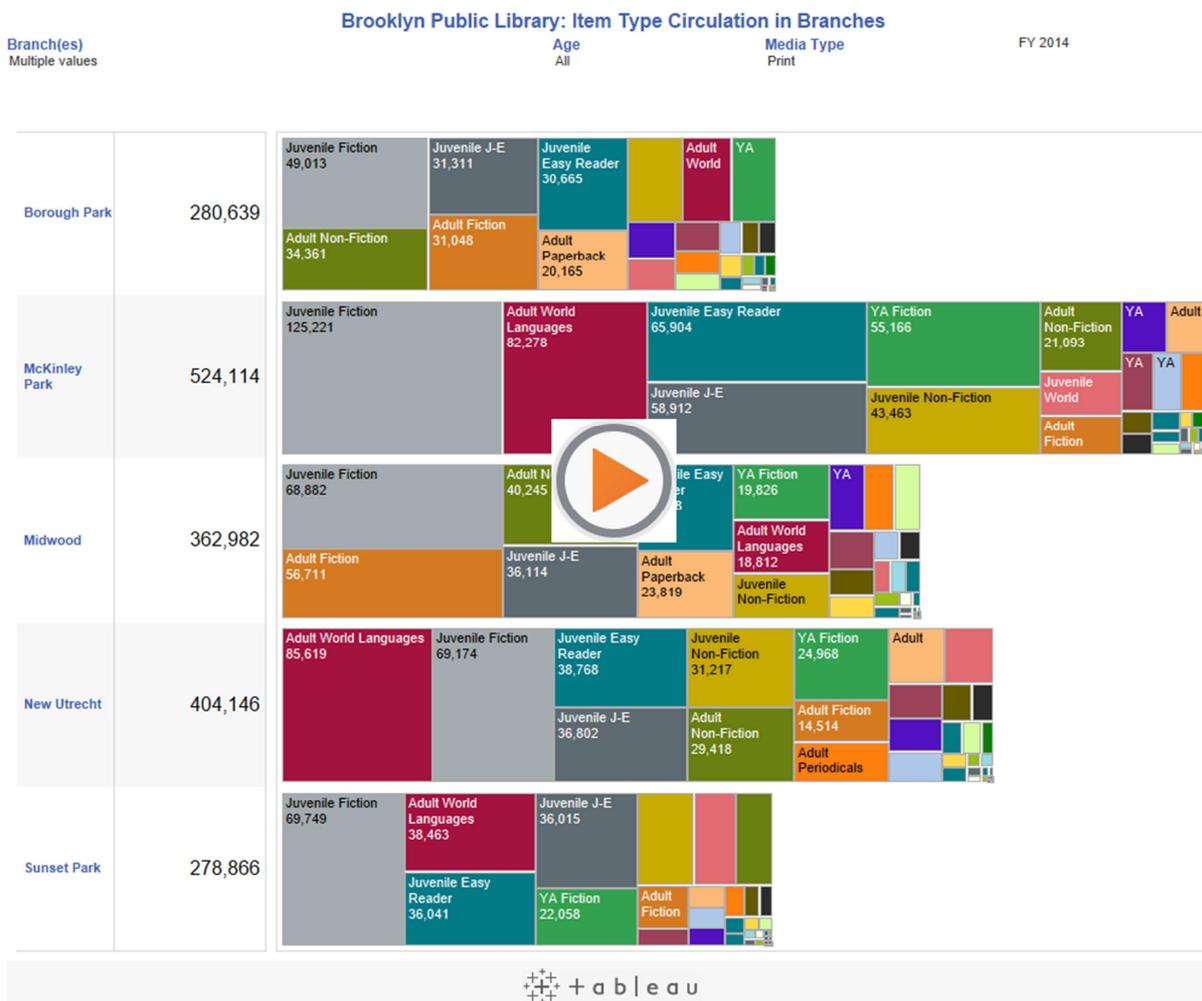


Figure 4. Tableau presentation of the BPL data analysis. (source: “Brooklyn Public Library: Item Type Circulation in Branches” (available online: <https://laobingkaisuo.com/tableau-spreadsheet/tableau-google-spreadsheet/> site visited on 06.11.2018), Page 8.).

Although the majority of the PBL data is conventional statistical data (such as visits, loans etc.), the new kind of representation helps the staff to understand the information better and implement it in a relevant way.

The following advantages for the library were derived: - better allocation of staff resources: five existing external

consultant positions were replaced with a new member of staff and an analyst assigned differently. - the duration of the reporting process for the monthly report was reduced by two weeks. - resolving accounting issues within the scope of checking user fines. - clear and easily comprehensible reports are now compiled, which are distributed to all members of staff.

In general, more ad hoc analyses are conducted in the BPL today that are based on both local data sources and the BPL Data Warehouse. This helps established a cultural of change within the library and enables the autonomous discovery of new possible courses of action for the staff [24, 25].

2.6. Joint Big Data Initiative Between Ten US Libraries

A ground-breaking example application is the joint big data project between ten public libraries from all over the USA, the Institute of Museum and Library Services (IMLS) and CIVICTechnologies, a software provider for data analyses. According to their own statements, it is supposedly the first true big data project in history. Ten medium-sized to large libraries are cooperating on this project to use the possibilities of big data analyses jointly. The ten participating project libraries operate in an area with a population of 7.8 million, over half of whom have a library card (4 million users or 52% of the overall population). In 2014 a total of 67.4 million (printed and digital) media were borrowed at

these libraries. The goal of this project was to get to know the customers and non-customers and their needs precisely, and gear the services towards them. The core questions in this context:

- who are the most active users of these public libraries, i.e. the “core customers”?
 - what are these core customers’ lifestyle habits?
 - what are their interests, preferences and behaviours?
 - what can we learn from people who frequently borrow or use media to keep them happy and help the libraries develop?
- Extensive public statistics and data at regional and national level (“census data”) were used for the data analysis and then linked to the usage data from the individual libraries. As a consequence, a better user experience should be created, (new) popular services and programmes offered and better strategic plans developed.

It is apparent that communities and library user groups are complex structures. The transfer and representation of this data on GIS-compatible maps for market segmentation illustrates this complexity in a striking way.

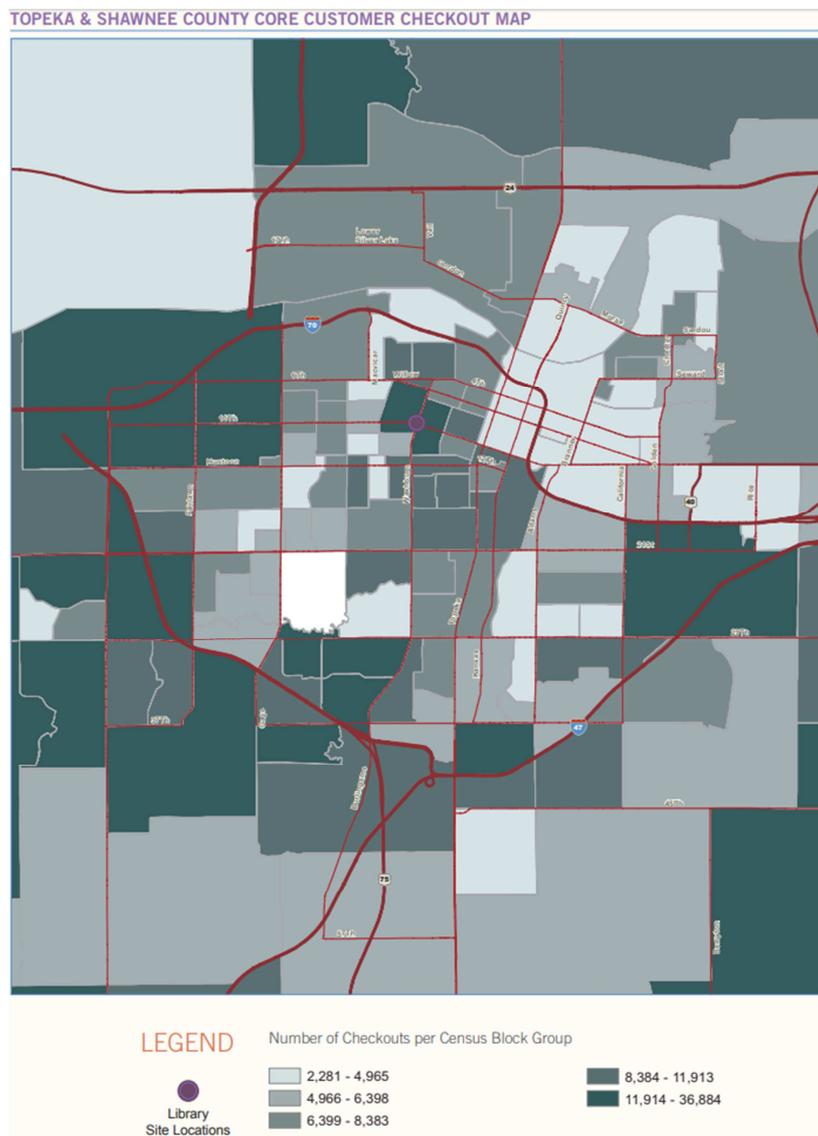


Figure 5. Number of loans according to census blocks. (source available online: <http://civicttechnologies.com/wp-content/themes/civicttechnologies/documents/core-customer-intelligence.pdf> site visited on 06.11.2018, Page 62).

The users of Las Vegas-Clark Country Library, for instance, come from 21 different types of household. This information alone ensured that the offline and online resources could be tailored in such a way as to address particular segments of the overall population specifically. [26]

3. Conclusions

In recent years, the interest among libraries in the topic of big data has markedly increased. Until now, the number of projects and initiatives in which libraries and/or information specialists have been involved in this field has been negligible, however. Nevertheless, the case studies presented here clearly demonstrate the enormous potential that big data harbours for libraries. The application possibilities range from improving the libraries' own services or creating completely new services to marketing measures, developing new data formats for library data for a more effective exchange, providing library data for data analysis for scientists, data standardisation, data modelling, library data visualisation and user behaviour studies. In principle, data analyses are thus possible at both micro and macro level.

And libraries do not just have to restrict themselves to their own data. They can also process and analyse external data sources. Meanwhile, numerous data sources from third parties are available that enable interesting services for the users to be extracted from them.

Limiting factors are currently still the lack of qualified skilled staff, the frequent lack of infrastructure, technical challenges (e.g. data formats and tools), data protection and funding problems.

References

- [1] J. Bailey (September 2015), "Top 4 ways to apply big data in higher education," in *Information Age*, (<http://www.information-age.com/top-4-ways-apply-big-data-higher-education-123460114/> accessed on 06.11.2018).
- [2] "Big Data," in Wikipedia. Die freie Enzyklopädie. (available online: https://de.wikipedia.org/wiki/Big_Data site visited on 07.11.2018).
- [3] W. Christl (November 2014), "Kommerzielle digitale Überwachung im Alltag. Erfassung, Verknüpfung und Verwertung persönlicher Daten im Zeitalter von Big Data: Internationale Trends, Risiken und Herausforderungen anhand ausgewählter Problemfelder und Beispiele," in Studie im Auftrag der Bundesarbeitskammer. Wien (available online: http://crackedlabs.org/dl/Studie_Digitale_Ueberwachung.pdf site visited on 07.11.2018).
- [4] Education, Communications and Outreach Group Cern (2017). "LHC: the guide" CERN-Brochure-2017-002-Eng (available online: <http://cds.cern.ch/record/2255762/files/CERN-Brochure-2017-002-Eng.pdf> site visited on 06.11.2018).
- [5] K. Metzler, D. A. Kim, N. Allum and A. Denman (2016), "Who Is Doing Computational Social Science? Trends in Big Data Research (White Paper)," London (available online: <https://us.sagepub.com/sites/default/files/compscsci.pdf> site visited on 06.11.2018). p. 8 ff.
- [6] A. Regalado (May 2013), "The Data Made Me Do It. The next frontier for big data is the individual," in *MIT Technology Review*, (available online: <https://www.technologyreview.com/s/514346/the-data-made-me-do-it/> site visited on 06.11.2018).
- [7] W. Bredemeier (February 2017), "Open Password (Interview mit R. Mumenthaler)," in *Password*, (available online: http://www.password-online.de/?wysija-page=1&controller=email&action=view&email_id=219 site visited on 06.11.2018).
- [8] M. Humbel (March 2017), "Die Umsetzung von Open Data an Wissenschaftlichen Bibliotheken der Schweiz. Eine qualitative Untersuchung," in *Churer Schriften zur Informationswissenschaft*, Schrift 86, (available online: https://www.htwchur.ch/fileadmin/htw_chur/angewandte_zukunftstechnologien/SII/churer_schriften/CS186-Umsetzung_Open_Data_an_wissenschaftlichen_Bibliotheken_der_Schweiz.pdf site visited on 03.10.2018).
- [9] M. Humbel (March 2017), "Die Umsetzung von Open Data an Wissenschaftlichen Bibliotheken der Schweiz. Eine qualitative Untersuchung (Interview mit M. Ehrismann, M. Hotea and R. D. Wanger)," in *Churer Schriften zur Informationswissenschaft*, Schrift 86, (available online: https://www.htwchur.ch/fileadmin/htw_chur/angewandte_zukunftstechnologien/SII/churer_schriften/CS186-Umsetzung_Open_Data_an_wissenschaftlichen_Bibliotheken_der_Schweiz.pdf site visited on 03.10.2018), 74ff. and 135.
- [10] T. Aaron (November 2016), "5 reasons why library analytics is on the rise (Blogbeitrag)," in *Musings about librarianship*, (online available: <https://musingsaboutlibrarianship.blogspot.com/2016/11/5-reasons-why-library-analytics-is-on.html> site visited on 06.11.2018).
- [11] A. McGill (August 2016), "Can Twitter Fit Inside the Library of Congress?" in *The Atlantic*, (available online: <https://www.theatlantic.com/technology/archive/2016/08/can-twitter-fit-inside-the-library-of-congress/494339/> site visited on 06.11.2018).
- [12] A. Foster (April 2013), "Add value or die. The fate of corporate information services. The Business Information Survey 2013" in *BIR*, (available online: <http://journals.sagepub.com/doi/full/10.1177/0266382113484222> site visited on 06.11.2018).
- [13] Alia (2016), "Website of the National 2016 Conference, 29 August-2 September 2016 Adelaide: Engage Create Lead" (available online: <https://library.alia.org.au/big-data-small-library-0> site visited on 25.10.2018)
- [14] V. Johnson (2016), "Big Data, Small Library," *Deakin* (available online: https://library.alia.org.au/file/424/download?token=rx_SWTG_Z site visited on 06.11.2018).
- [15] V. Johnson (2016), "Big Data, Small Library," *Deakin* (available online: <https://library.alia.org.au/big-data-small-library-0> accessed on 06.11.2018)

- [16] University of Michigan Library (2018), "Open Access Bibliographic Records Available for Download and Use" (available online: <https://www.lib.umich.edu/open-access-bibliographic-records> site visited on 06.11.2018).
- [17] Harvard Library (2012), "Harvard Library Open Metadata", (available online: <https://emeritus.library.harvard.edu/open-metadata#Harvard-Library-Bibliographic-Dataset> site visited on: 06.11.2018).
- [18] D. Black (April 2012), "I'm Tired of Hearing about Big Data" in The Black Liszt, (available online: <http://www.blackliszt.com/2012/04/im-tired-of-hearing-about-big-data.html> site visited on: 06.11.2018).
- [19] Q. Hardy (April 2012), "Harvard Releases Big Data for Books" in The New York Times (available online: https://bits.blogs.nytimes.com/2012/04/24/harvard-releases-big-data-for-books/?_r=0<https://www.ethz.ch/de.html> site visited on 06.11.2018).
- [20] Jisc (2018), "Jisc Website", (available online: <https://www.jisc.ac.uk/> site visited on 06.11.2018).
- [21] Higher Education Statistics Agency (2018): "HESA Website" (available online: <https://www.hesa.ac.uk/services/heidi-plus> site visited on 06.11.2018).
- [22] S. Burke (October 2016), "Library + Data + Labs = what exactly?" in Jisc, (available online: <https://libraryservices.jiscinvolve.org/wp/2016/10/library-data-labs-project/> site visited on 06.11.2018).
- [23] Tableau Software (2018): "Tableau Website", (available online: <https://www.tableau.com/> site visited on 06.11.2018).
- [24] C. O'Maley Voliva (April 2015), "Data Visualization for Public Libraries," in Public Libraries Online, (available online: <http://publiclibrariesonline.org/2015/04/data-visualization-for-public-libraries/> site visited on 06.11.2018).
- [25] Tableau (2018), "Brooklyn Public Library saves time, money and headcount with Tableau", (available online: <https://www.tableau.com/solutions/customer/brooklyn-public-library-saves-time-money-and-headcount-tableau> site visited on 06.11.2018).
- [26] M. Futterman, D. P. Milam (March 2016), "Core Customer Intelligence: Public Library Reach, Relevance and Resilience," in CivicTechnologies, (available online: <http://civictechnologies.com/core-customer-intelligence> site visited on 06.11.2018).