Qualitative and Quantitative Methods in Libraries (QQML) 5: 287-300, 2016

Project BLIIPS: Making the Physical Public Library more Intelligent through Artificial Intelligence

Martijn van Otterlo, PhD

Amsterdam Center for Business Analytics (ACBA), Data Science Alkmaar (DSA), KIN Research/FEWEB and Amsterdam Academic Alliance Data Science (AAA-DS)
Vrije Universiteit Amsterdam, The Netherlands, http://martijnvanotterlo.nl

Abstract: This paper reports on the underlying motivations of the ongoing project BLIIPS to make public libraries more intelligent by data-driven optimization. Four key developments are described which drive the approach: data gathering in physical spaces using sensors, retail approaches in current public libraries, intelligent algorithms for data analysis and experimentation with data-driven strategies. I highlight these developments to arrive at the final target: data-driven, experimental, physical, public library innovation.

Keywords: digitalization, artificial intelligence, library use, data analytics, intelligent algorithms, sensors, libraryness, physical-digital continuum, public library.

1. Introduction

Artificial intelligence (AI) (Nilsson, 2010) is spreading rapidly in our society, making our Google interactions smarter, making facial recognition on social networks feasible, or creating possibilities for human-aware robots. Even though information-rich public libraries seem to be ideal environments for the application of such smart, digital technologies, as we will argue in this paper, efforts have been limited. Measuring the activities in the (physical) library and analysing the obtained data intelligently holds great potential to aid in the design of the physical space, the optimization of various information services, and the improvement of satisfaction of both patrons and librarians. In fact, such data analysis may also give rise to new (and fun) ways of using the physical library and its collection in the library of the future.

Writing about *the future of the library* more than fifty years ago, Licklider (1965:33) noted: *"By the year 2000, information and knowledge may be as important as mobility. We are assuming that the average man of that year may*

Received: 24.2.2016 Accepted: 21.4.2016 ISSN 2241-1925

make a capital investment in an "intermedium" or console – his intellectual Ford or Cadillac – comparable to the investment he now makes in an automobile, or that he will rent one from a public utility that handles information processing as Consolidated Edison handles electric power." Today, this intermedium is the smartphone which most people carry everywhere they go, and which provides access to a dazzling amount of information. Licklider predicted that people would be more directly connected to the total "body of knowledge" instead of through individual books, using a so-called "procognitive utility net", which shares many features with our modern internet. It also aligns with one of the key developments in modern libraries: people increasingly get their information from sources outside the library, mainly from private entities such as Google. Bypasssing the library, in some (digital) way, is not new[1]: many decades ago Vanevar Bush introduced his *Memex*, and even longer ago H.G. Wells described his *World Brain*. Both ideas were centered around bypassing (at least) the physical library, making use of "remote access" to knowledge with technology, but many other developments can be expected (Noh, 2015).

Besides being bypassed by digital technologies, the public library is in transition, influenced by many opposing forces. The internet, commercial forces and a decrease of pure reading activities (and book loans) are contrasted with the many societal functions that are expected, the central, "third place" property of public libraries, and the fact that many people still rely on the public library to get access to "information" and to obtain "21st century skills" in many forms. The key aspect of the public library is the *physical* building with its *physical* book collection. Digital technologies are used simultaneously, for example e-books, or web-based catalogues, but the physical nature of the public library remains its greatest asset. *"One might infer that once the books are no longer in analog format, the need for library spaces will go away. That inference turns out to be wrong."* (Palfrey, 2015:66). The public library is a physical place, where patrons frequently go to, to borrow books, to meet, and to be informed. However, despite many surveys, studies, and innovative approaches in the modern public library, real-time insight into the activities of patrons in the physical library space with its physical collection, and insight in optimization opportunities for various library services, are severely lacking.

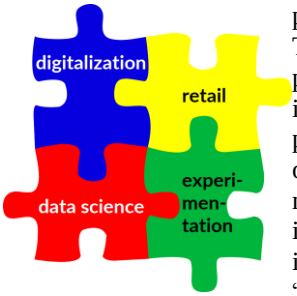
In this paper I argue for the use of *data science* (AI) in the physical, public library in the context of a just started project: BLIIPS. My work is situated in a larger effort to understand, utilize and even optimize (physical) aspects of the public library of the future, by digital means. This paper consists of a small, but important, fragment of a much larger literature study on the interplay between digital and physical aspects of public libraries. Practically the project is aimed at developing several data-oriented technologies to aid patrons and librarians in various information activities, in a designated collection of libraries in the Netherlands. Current developments in "sensors" drive this. Overall, I named it BLIIPS, which stands for *"Books and Libraries: Intelligence and Interaction through Puzzle- and Skinnerboxes"* and this paper outlines some of its key

1 Space is limited here, but I refer the interested reader to Wright (2014).

motivations. Based on literature and initial investigation in the participating libraries, I identify four interlocking motivations that, together, induce a research program for experimental data science in the physical, public library.

2. Four key developments driving data-driven public libraries

BLIIPS is associated with the physical (main) library in the Dutch town Alkmaar. Its core aim is to use smartphones and related technology to sense and record how patrons interact with the physical library and its books, and to aid a patron in navigating to "possibly interesting" books. The results of the project will be beneficial for patrons, library management and science. The latter is about a general, data-oriented understanding of the physical (public) library and its books in the context of patron activities. For librarians the data-driven management of the library and its services is an interesting target, whereas for patrons we look at issues such as how can one make a visit to the library "better" or possibly a fun "experience". All three will be pursued simultaneously: for example, if smartphone technology can be utilized to localize a patron in the physical library, both the patron and the management will benefit from a context-based, intelligent, persuasive, personalized book recommendation.



My vision on the future of the physical library is motivated by four key, interlocking developments, see the figure. Although they are to be seen as four individual (although related) developments, the order in which I discuss them in the following section does matter in terms of how together they construct the main rationale for a data-driven public library of the future. The "digitalization" piece deals with obtaining data from physical interactions (*data gathering*). The "retail" piece is about a recent trend in (Dutch) libraries to approach public library management using marketing tools, and generates *goals* for measuring in the first place. The "data science" piece then deals with statistical algorithms (*automated analysis*) making sense of gathered data. Finally, the "experimentation" piece is the iterative, interactive and repeated application of automated analysis to, basically, find out "how to *optimize* the library using data". One end goal of this research, the complete puzzle, is to establish a "library laboratory" in which data is purposely, and repeatedly, gathered and analysed automatically, generating an experimental loop to optimize the library.

2.1. Towards digitalizing all physical (inter)actions in the library

The first puzzle piece deals with digitalization of the physical space. Umberto Eco's novel "The name of the Rose" (1980) contains, among many things, a story about a *book search*, in which a forbidden book is to be found in a physically complicated library, with an evil librarian obstructing the search. Eco listed 19 possible (physical) obstructions in the modern library in his foreword to Hoefer's (2005) book, including unfindable books and incomplete catalogues. According to recent surveys, core problems of patrons in the

physical library are navigation and finding books and topics, despite the many advances in signage, catalogues and space planning (Edwards, 2009 ch.4). In addition, physical libraries and books require physically navigating, walking, and picking up books, and for library management it requires many other physical actions such as sorting, transporting, placing, relocating, and so on.

How different this is for digital collections and the general internet. Most such services will allow a person to enter a couple of words (search query) that somehow characterize things being searched, and typically one instantly obtains a ranked list of items, without "navigating" to them, ready to be investigated immediately. Furthermore, digital collections can be ordered on-the-fly based on any contextual information, and many patrons can have simultaneous access to the same text. In addition, digital services can offer "interesting" items not directly searched for, but which may be of interest, based on (inferred) interests.

To provide the "digital" kind of user experience in the physical library, one needs to *digitalize* the interactions with patrons and books using modern technology. Smart technology could *sense* patrons and books and based on this, an *intelligent algorithm* could find out the quickest route from the patron's current location to a book. An early example of such a *location-aware mobile library service* is *SmartLibrary* by Aittola et al. (2003) which guides the patron to required books using a map displayed on a digital device. Another early approach by Satphaty and Anijo (2006) uses RFID chips to measure proximity (between patrons and books) for similar book search tasks. Both used PDAs which are precursors to the modern smartphone. More recently other opportunities arise to measure and help in the physical library with technologies like QR-codes and location-aware technologies such as GPS, WiFi and beacons. Walsh (2010) lists several new possibilities such as linking books and electronic resources, and associating video help with physical books and objects. The rise of intelligent image and video analysis will also give new opportunities in smartphones such as *virtual reality*, and *augmented reality* – which allows for additional information being projected onto images, or which can recognize books and objects using the smartphone camera – as described by Hahn (2012). Current developments in both hardware (sensors, smartphones, cameras) and intelligent software to "interpret" rich data from these sensors, will lift our potential to sense and record interactions in the physical library considerably (see also "library 4.0." by Noh, 2015). Mobile technology can also be used to aid in traditional methodologies such as observation (Thompson, 2015).

Several other studies exist to *augment* the physical library in such a way that patrons can be tracked in the library, books can be located, objects can become "smart" and interactions between patrons, books and the library space can be sensed and recorded. However, the literature indicates that so far all studies are limited to very specific technologies and/or small user groups and quite often developed as either a proof-of-concept or as a means to test a particular hypothesis (e.g. about search or browsing behavior). There is no widespread use of any of the developed systems yet, even though it would be highly desirable.

Overall, the rise of smartphones, sensors, and intelligent software holds great promise to "free the library from (some) physical constraints". One interesting

dot on the horizon could be the conceptual, future library project named *"the learning jungle"* as envisioned by architects (Dijkstra and Hilgefort, 2010) in which the collection and its physical layout in the library space are solely determined by the patrons themselves and where books and patrons interactively make up new orderings in the physical space. This would make the physical space much more similar to digital libraries (see also van Otterlo, 2015)).

2.2. Retail thinking and customer journeys in the physical public library

The key aspect of the second puzzle piece is that many public libraries are now being transformed with the help of marketing tools. The so-called *retail*[2] *approach* basically treats the library like a *store* and its patrons as *customers*. This has profound implications for thinking about collections, patrons, and the core functions of the library. Marketing is not new in the (public) library, and includes well-known aspects such as strategic plans (vision and mission), promotion (messages, campaigns) and the way services are provided in general (Garoufallou et al. 2013). New social media (Facebook, Twitter) enable many new forms of personalized, promotions. An important aspect is market research, which gets much attention in (the retail approach to) modern public libraries, and can affect quite significantly the "looks" and "feel" of the physical library.

The retail approach (Stanley, 2003) utilizes existing (empirical) knowledge about success factors of retail stores and applies them to library management. Typical expected benefits are "better" patron *experience* (e.g. navigation, book finding, enjoyment), more book loans, better use of space and collection, more time spent by patrons in the library, more returning patrons and so on. Retail methods also come with implementation strategies to successfully change the (human) organization, since for many librarians, treating the patron as a mere customer may not be a natural thing to do (although for customers it may very well be). Fundamental aspects are the "customer centric" stance and the emphasis on *self-service*: the customer must be able to do much as possible – e.g. all activities for borrowing – independently, which changes interaction with the staff. Also important is the emphasis on "sales": trying to "sell" the books and connect customers to other items (books) they might like (persuasion).

For patrons, the most visible aspects of retail strategies are the changes to shelves, classification systems, collection content, and especially the physical layout. A typical "retail library" will have clusters such as "kids 0-8", "young adults 12+", "literature and culture" and "mind and body" to cluster topics and books. This resembles general book *stores* and deviates from typical library classification systems such as UDC. The resulting clusters also form the basis for a detailed spatial layout in which, for example, typical topics for young mothers are "close" to the topic "kids 0-8", and preferably "back in" the library. The display of books is affected in many ways: i) shelves are altered such that many books are displayed with their cover facing towards customers, and in

2 https://www.debibliotheekformule.nl/ is a typical (Dutch) example, used in the BLIIPS library too, which promises more book loans and more returning (paying) customers, among others.

total less shelves are present, ii) power displays, and "end-caps" display books that are "promoted" in a particular period, for example because they are new arrivals, new acquisitions, or because they are grouped by a particular, current theme, and iii) signage and imagery is placed in clusters to draw attention. All these aspects are cleverly combined to maximize space and to optimize the "sales" of the library. Typically, this plan of the library starts at the entrance where various power displays are present to persuade people to enter the library, and to immediately show interesting content.

Retail strategies employ a body of knowledge about stores and empirical studies on customer behavior and several tools exist for analysis (and optimization). Framed more familiarly as "service design" (Marquez and Downey, 2015), to optimize the library's services one can look at co-creation, blueprints, and many other (usability design) techniques. Here we focus on the popular concept of *customer journeys* (Nenonen et al. 2008), which are *process- and experience oriented*, and deal with the *interaction* between a customer and a service (in this case the library), with *the aim to understand what customers typically do during the service process.* The interaction is structured using *touchpoints* which are those occasions where a particular type of interaction takes place. For example, this could be "signing up for a membership", "notifying a problem", or "borrowing a book". Customer journeys can be constructed employing people's mental models, scenarios, and *personas,* which are "typical" customers. Personas[3] are used to generalize over individual customers, since many of them will have very similar interactions with a service.

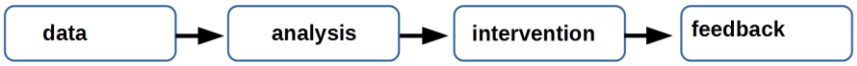
Now, even though library marketeers, especially in "retail" libraries, may know much about their customers, their "market segment", their book loan history, and personal information, most libraries do not know much about the actual *use* of the *physical library*. Returning books and registering new loans are registered touchpoints, but what happens in between goes by undocumented. Building on the developments in the previous section, I propose to *enrich* the existing customer journeys studied by marketeers with those from the *physical journeys* through the library, with numerous additional touchpoints such as visiting specific locations, interacting with specific books, and querying at various locations in the library. Personas and customer journeys could now take into account *all* information about how people use the *physical* library with all (electronic and nonelectronic) services.

2.3. Data, data science and artificial intelligence

The third piece of the puzzle is about data and algorithms. A century ago the *scientific* way of doing library planning was about detailed calculations of simple shelf widths, room sizes and lighting conditions (Tilton 1915). The last decades have shown much more involved strategies for *library design* (Edwards, 2009), utilizing behavioral studies and retail approaches, and now we

3 Mosaic (http://www.experian.nl/mosaic/) is often used as a general segmentation of all citizens of the Netherlands into groups which can be used for marketing purposes.

have entered the age of *big data* (McAfee and Brynjolfsson, 2012) which promises to solve any design or optimization problem using *data* and *algorithms*. *Data science* is a recent term for many similar things like data analysis, statistics, *datamining*, AI, *machine learning* and related techniques.



Typical data science efforts involve roughly four steps. Obviously the first is obtaining data; in the previous I have already indicated some difficulties of obtaining it in physical situations. The second (analysis) step is crucial, since it involves *intelligent algorithms* working on the data to find (statistical) *patterns* or to compute *generalized* statements about the data. A most common usage is *profiling* (van Otterlo, 2013) in which data is *grouped* in such a way that the groups satisfy useful properties. This way, customers can be profiled (based on their data) as "likely to read", "high income", or "lots of free time". Such profiles can then be used in the third step, to *intervene* based on the analysis. For example, Amazon can do this in their webshop by offering me a deal on a book, or by showing me "related" books that I might also like. A key issue here is that they use the outcome of the analysis of the data to *act* upon the information obtained. This will likely induce the fourth step: feedback. By either "liking" certain content, or by actually buying the suggested book, a customer provides feedback on the intervention based on the analysis of the data. Many variants of this four-step process exist: if the data is mere demographic data, the analysis is much different than if the data corresponds to behavioral data (e.g. click data on a website) rendering the analysis a form of *activity recognition* (Yang, 2009).

Data in the library comes in many forms, and so do analyses. For electronic services and resources, more work has been done on the *analytics*[4] (Showers, 2015). However, for data science in the physical, public library, things are only just starting: data gathering is often labour-intensive, and automated analyses are still limited. Khoo et. al. (2012) extensively survey ethnographic techniques that have been used in libraries, with general types of methods such as observations, interviews, fieldwork, focus groups and cultural probes, and many examples of studies on specific user groups or library services, on space, wayfinding, seating occupation and so on. Mandel (2013) used (unobtrusive) observations, extensive interviews and expert consultations to study wayfinding behaviors in public libraries. Three weeks of data for a specific purpose (hypotheses) was gathered and analysed using graphical information systems (GIS) to visualize behavioral patterns of patrons. Analysis of these patterns gave rise to generalized concepts such as "high-traffic areas" and "popular routes". Both data gathering and analysis relied on much human labour, and so far, intervention and feedback steps have not been implemented. Given and Archibald (2015) used mobile devices for easier (but human-involved) data

4 Examples: UK http://jisclamp.mimas.ac.uk/ and US https://osc.hul.harvard.edu/liblab/

gathering, and used GIS too for analysis, which resulted in insights about things like a "lack of use of physical collection" and "quiet and noisy areas". Khoo et al. (2016) studied in an academic library the change from individual study to emergent-technology-supported group study. They used a mixed approach of 2 qualitative, large surveys and quantitative data about seating use. R-Moreno et al. (2014) report on a more automated system in a public library in which data is gathered using RFID sensors (i.e. the patron carries a chip) in gates between rooms, and where the system uses AI *planning* to generate route descriptions for individual patrons. Information about the route is displayed on several screens in the library. An interesting fact is that this system semi-automatically implements all four data science steps, and can report on room occupancy and the time a patron spent in the library semi-realtime.
Overall, this is just a sample of the approaches in the literature, but the findings are representative. So far, most previous research focuses on the first two steps (gathering and analysis) and this happens predominantly qualitatively and labour-intensively.

2.4. Experimentation and optimization

Finally our fourth and last piece of the puzzle is *optimization.* Optimizing behaviors has always been a key issue in experimental psychology, with well-known scholars such as Watson, Pavlov, Thorndike and Skinner, who pioneered with cleverly designed experiments to study the fundamentals of behavior. *Skinnerboxes* are small experimental setups where animals (mice, pigeons) are placed under controlled conditions. Usually there is some desired coupling of *stimuli* (light, sounds) and *actions* (pressing a lever) that the animals need to *learn*, motivated accordingly using food (*rewards*). Another tool is the *puzzlebox*, used in Thorndike's work, containing several mechanical parts, from which an animal needs to *escape*, motivated by food outside the box. Both tools focus at *shaping* behaviors, and much is known about various learning schemes and conditions for learning due to these experiments. A key issue is the *experimentation* itself, in which one tries to generalize and understand, possibly based on prior hypotheses, but also to *try out* things to *see what happens*.
The acronym BLIIPS contains both Puzzle- as Skinnerbox, since their modern, data science variant is a subfield of AI called *reinforcement learning (RL)* (Wiering and van Otterlo, 2012). RL is a suite of algorithms which targets the learning and optimization of *strategies*. A recent, very impressive achievement of RL is beating of the human champion in the difficult board game Go[5], but RL has show success in various other tasks, including many in robotics. RL typically implements all four steps of data science: gathering, analysis, intervention and feedback, but a crucial addition is what I call "closing the loop". Imagine a robot learning how to grab a book by just trying out many times. At first it might not be able to even approach it well, later it may be possible to touch it, even later it may have some crude grasp but drops the book often, and finally it can master the whole sequence of actions required to

5 https://deepmind.com/alpha-go.html

approach and grab the book. In terms of data, the robot gathers and analyses what it is currently doing, and based on that it may act (intervene) and gets evaluated in terms of how successful the grab was (feedback). Closing the loop now means using this feedback to *adapt* the behavior somewhat: make it more likely to do something similar again if the feedback is positive, and less likely if the feedback is negative. The new behavior results in new data, which can be analyzed, and so on. Closing the loop means having multiple data science sequences lined up, and where feedback is used to *steer* the process. Feedback can be anything that provides a *measure* on *how well* things are going.

Automated experimental loops are becoming more and more a reality (van Otterlo, 2014), especially in digital domains. Google analytics offers many opportunities for experimenting with subtle variants of websites or apps, so-called *A/B testing*[6] which has been applied in library settings as well (Young, 2014). Facebook even has its own programming language (Bakshy et. al., 2014) to set up huge *field experiments* with millions of Facebook users, with infamous examples in *mood*[7] *manipulation* and *elections*[8]. And of course, retail websites such as Amazon do it all the time too: gathering and analysing our data, making us an offer (intervention) and seeing our feedback (did we buy it?). Such techniques even make it possible to experiment with *pricing* mechanisms: how to optimize prices for products and customers individually, such that the global (estimated) sales result is best?

3. Synthesis: towards a physical library laboratory

So far, I have described four developments which I think together form important building blocks of the physical, public library of the future. I have first identified a need for, and a general lack of, information about how patrons use the *physical* library and how these interactions could be *digitalized* using smartphones and smart sensors. Then I described so-called *retail* strategies in which libraries are treated as stores, patrons as customers and where much effort is put into *experience*, *sales* and *customer research.* The retail setup provides an interesting setting with very clear goals for optimization of the "store". I then invoked *data science* as a general tool to automatically obtain, analyse and use data for various goals: activity recognition, navigation assistance, location-based recommendations, recognizing "typical" customers, and so on. Finally, I described reinforcement learning as a general toolbox for *optimization* of strategies, which can represent various library design decisions but also the construction of various library *services*. Together, these components define a library situation in which not only much more is known about the activities in the physical library, but in which various aspects of this physical library can be optimized using data. These new opportunities for *data-driven innovation in the*

6 https://developers.google.com/analytics/solutions/experiments

7 http://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/

8 http://www.motherjones.com/politics/2014/10/can-voting-facebook-button-improve-voter-turnout

(physical) public library are, so far, severely under-explored. My project BLIIPS is a serious attempt to make progress in this direction.

On a more general level, several authors seem to have similar desires for library management. For example, Given and Archibald (2015:102) write: *"...librarians need tools that allow for detailed understanding of various options for arranging stacks, reading areas, and workstations".* Khoo et al. (2016:56) mention *"...an action research approach, which gathers and analyzes data, generates recommendations for intervention, and builds theory, on an ongoing basis.".* My description in this paper is novel in the sense that it defines a much more general setting but at the same time links it to much more concrete developments, most importantly in data science.

As another motivating example, consider the (retail) study by Stern (2010) in which observations, interviews and video were used to obtain data about user activities and where the main goal was to obtain an understanding of how public libraries can better service and educate their visitors. Three interesting findings were that more than half of the patrons stayed less than 10 minutes per visit, two thirds of patrons came unprepared (i.e. without clear goals) and the majority came alone. An interesting intervention was done with their 400+ magazines, which circulated rarely and were difficult to browse. By displaying them differently (facing outwards more) and placing them elsewhere (for example near working spaces so that people waiting for a space could browse them) magazine circulation increased by 40 percent.

One can imagine that many such simple or complex interventions could be tried or verified, almost daily, if only data would be collected all the time to see how decisions turn out. I especially call for more *long-term, continuously measuring* investigations of data science in the public library, to measure, analyse, intervene and incorporate rich feedback in order to improve services and design of the library. Equally importantly, I am arguing for more, and better grounded, *experimentation* in the library, using data. Lastly, I am arguing for more *automated experimentation*, using *algorithms.* The public library, being a flexible, non-profit organization, seems an ideal "laboratory" for all of this. As Palfrey (2015:213) writes: *"An innovative librarian from Finland, Kari Lamsa, said it well in an interview: 'Libraries are not so serious places. We should not be too afraid of mistakes. We are not hospitals. We cannot kill people here. We can make mistakes and nobody will die. We can try and test and try and test all the time.'"* Data-driven strategies will allow for *evidence-based* decisions: based on some hypothesis one could see whether one strategy "X" beats strategy "Y". In addition it would be possible to test *multiple* different ones, and see in the resulting data, even *real-time*, which is best. In fact, data-driven strategies also allow for *bottom-up* induction of possible *new* strategies and support informed testing of them. We need a comprehensive approach, and to be able to experiment with services and the library space and collection design.

4. Outlook

Although library analytics does exist in some forms, I have argued that most work is either aimed at contexts that are already digital, or aimed at localized

experiments, and certainly almost always not automated. I envision a data-driven future of the physical, public library in which data gathering, analysis, interventions and feedback utilization is done by intelligent algorithms optimizing various library services. Project BLIIPS aims at starting a new, general framework in which many data-oriented methods can be investigated, based on AI and reinforcement learning. Many important challenges, and opportunities, lie ahead. Four key challenges concern data gathering, optimization, the use of domain knowledge, and the ethics of manipulation algorithms, and I conclude by briefly discussing each of them.

Concerning the measurement of data, immediate challenges are technical in nature. Digitalizing physical (inter)actions can sometimes be seen as an attempt to automate various traditional techniques such as observations and fieldwork. Since automated video-analysis has matured but not to the level that general human activities can be recognized, still a lot of work is needed to cover *all* interactions in the physical library. In BLIIPS we utilize current sensor and smartphone technology, but the setup should be general enough to extend when new technology becomes available. The aim should also be to *track* patrons, personnel and objects (books, desks) throughout the library, such that the exact physical order in the library becomes more flexible. This way, it gives the personnel more opportunities to keep track of the collection, but it also gives the means for patrons to be guided by their smartphones to interesting books. The most important goal is to measure *continuously*, with many users, and do analyses in *real-time*. This requires intelligent solutions for storage and computation as well. Another real challenge lies in constructing just the right circumstances in which people in the library are *assisted* in their existing interaction with the books and the library, or be *enticed* to pick up new routines.

This last remark hints to a second set of challenges, concerning *optimization*. One challenge is how to *manipulate, influence, nudge,* or *persuade* patrons to change their behavior in the library. Since much of the interaction can be done using smartphones, it is natural to think about *feedback*, *suggestions* or other types of information to *send* to the patron. However, one can also think of feedback *generated* by the patron, such as 5-star ratings, likes and dislikes, or general complaints. How to represent, interact and utilize these feedback mechanisms is part of the scientific research, but it may give rise to various ways to practically *optimize* a patron's effectiveness or satisfaction in navigating the library. Another optimization issue is coming up with *what* to optimize. In BLIIPS one of the main goals is to optimize the number of loans. This has been a main target for the retail strategy of the library too, and it is a highly general goal for which many interventions could be important. Experimenting with many types of interventions in services, library design or collection management could result in new, possibly unexpected, interventions that, somehow, increase the number of loans. An interesting possibility is to obtain experimental results that either confirm, or reject, the various commonsense (retail) design considerations on which public library design is based until now. A simple example could be the number of personas in any customer segmentation: it may very well be that experiments and data could indicate more or less detailed

segmentations are needed to optimize particular library services. Experimental procedures with data are key to optimize particular goals without having to know beforehand *how* to achieve them, just by *trying out*. This also means that quite general goals could be studied: imagine, for example, the library as one big puzzlebox, and imagine optimizing the opposite: can we *nudge* patrons into staying *in* the "library-puzzlebox" .... *longer?*

A third set of challenges comes from a need to provide insights into data, analysis and conclusions in intuitive terms. Library management and patrons are generally not interested in overly complete statistical figures and extreme details of data. This requires that feedback, generalized patterns, personas, "typical" user behaviors and required interventions should be represented and communicated in human-understandable format, which is quite a challenge for many modern data science methods. Visualizations such as GIS can help, but more research is needed into effective data visualizations.

A last (but certainly not least) issue when discussing behavioral manipulation in the library is *ethics*. Libraries have been defenders of privacy for a very long time, and *any* data collection in the library should be approached with care. Privacy is very important in the data-driven library setting I have described so far. Much has been written about societal consequences and legal considerations, but typically practical considerations make data collection in some occasions ethical or less so. Data science can really make the physical library more effective, and presumably more fun, but users should be given a choice, and be given transparent information about the trade-off between sharing information and obtaining benefits. Patron's borrowing history and reading habits, as part of our *intellectual privacy* (Richards, 2015), should be protected at all costs. In addition, if users are being thrown into *experiments* which are set up to change *their* behavior – however noble or practical the goals may be – we should first have a debate on that (van Otterlo, 2014), and users need to have a choice to "opt in" rather than "opt out". Finally, combining both issues, optimization in the library will enforce the library's role as knowledge *gatekeeper* (van Otterlo, 2015); another big issue to be studied in this context.

So, lots of challenges, and especially opportunities await for the data-driven physical, public library of the future. Hopefully, data science could shed more light on such intricate questions like: is the library a place, or a service, or a store, or an experience?

Acknowledgements

The author thanks the Amsterdam Academic Alliance Data Science (AAA-DS) consortium on data science and stichting Leenrecht for funding this research and Bibliotheek Kennemerwaard for their kind cooperation and support.

References

Aittola, M., Ryhanen, T. and Ojala, T. (2003) *SmartLibrary: Location-aware mobile library service*, in: *Proceedings of mobile HCI*, LNCS 2795, pp. 411-416.

Bakshy, E., Eckles, D. and Bernstein, M.S. (2014) *Designing and deploying online field experiments.* In Proceedings of the international conference on World wide web (WWW '14). pp. 283-292

Edwards, B. (2009) *Libraries and learning resource centres*, 2nd ed., Architectural press/Elsevier.
Dijkstra, R. And Hilgefort, J. (2010) *The Learning Jungle*, in: *The architecture of knowledge, the library of the future*, NAi Publishers.
Garoufallou, E., Siatri, R., Zafeiriou, G. and Balampanidou, E. (2013) *The use of marketing concepts in library services: a literature review*, Library Review, 62(4/5), pp. 312-334
Given, L.M. Archibald, H. (2015) *Visual traffic sweeps (VTS): A research method for mapping user activities in the library space*, Library & information science research, 37(2), pp. 100-108.
Hahn, J. (2012) *Mobile augmented reality applications for library services*, New library world, 113(9/10), pp.429 – 438.
Hoefer, C. (2005) *Libraries*, Schirmer/Mosel
Khoo, M., Rozaklis, L. and Hall, C. (2012) *A survey of the use of ethnographic methods in the study of libraries and library users*, Library & information science research, 34(2) pp. 82-91.
Khoo, M.J., Rozaklis, L., Hall, C. and Kusunoki, D. (2016) *"A Really Nice Spot": Evaluating Place, Space, and Technology in Academic Libraries*, College and research libraries, 77, pp. 51-70.
Licklider, J.C.R. (1965) *Libraries of the future*, MIT Press.
Mandel, L.H. (2013) *Finding their way: How public library users wayfind*, Library & Information Science Research, 35(4) pp. 264-271.
Marquez, J., Downey, A. (2015) *Service design: An introduction to a holistic assessment methodology of library services*, Journal of library user experience, 1(2).
McAfee, A. & Brynjolfsson, E. (2012) *Big Data: The Management Revolution*, Harvard Business Review, Oct.
Nenonen, S., Rasila, H., Junnonen, J.M. and Kärnä, S. (2008) *Customer journey - a method to investigate user experience*, in proc. of the Euro FM conference Manchester
Nilsson, N.J. (2010) *The quest for artificial intelligence*, Cambridge Universty Press
Noh, Y. (2015) *Imagining library 4.0: Creating a model for future libraries*, The journal of academic librarianship, 41(6), November 2015, pp. 786-797.
Palfrey, J. (2015) *Bibliotech*, Basic Books.
R-Moreno, M.D., Castaño, B. Barrero, D.F. and Hellín, A.M. (2014) *Efficient services management in libraries using AI and wireless techniques*, Expert systems with applications, 41(17), pp. 7904-7913.
Satpathy, L. and Anijo (2006) M. *RFID assistance system for faster book search in public libraries*, in: *Proceedings of CHI*, pp. 22-27
Showers, B. (ed., 2015) *Library analytics and metrics*, Facet Publishing, UK
Stanley, J. (2003) *Retail technology applications and their role in the modern library*, in: *Planning the modern library building*, McCabe, G.B. and Kennedy, J.R. (eds.), chapter 7, pp. 75-88.
Stern, V. (2010) *Why we borrow? Using retail strategies in public libraries*, ILA Reporter, june, pp 4-9.
Thompson, S. (2015) *Using mobile technology for studying library spaces*, Qualitative ans quantitative methods in libraries (QQML) 4, pp. 413-420.
Tilton, E.L. (1915) *Scientific library planning* in: *Library planning, bookstacks and shelving*, Snead & Co. Iron Works, pp. 113-117.
van Otterlo, M. (2013) *A machine learning perspective on profiling* in: *Privacy, due process and the computational turn*, Hildebrandt and de Vries (eds.), Routledge.
van Otterlo, M. (2014) *Automated experimentation in Walden 3.0*, in: Surveillance and society, 12(2).

van Otterlo, M. (2015) *The libraryness of calculative devices*, in: *Algorithmic life*, Amoore, L. and Piotukh, V. (eds.) Routledge.
Walsh, A. (2011) *Blurring the boundaries between our physical and electronic libraries: location-aware technologies, QR codes and RFID tags*, The electronic library, 29(4), pp. 429-437.
Wiering, M. A. and van Otterlo, M. (2012) *Reinforcement Learning*, Springer.
Wright, A. (2014) *Cataloging the world*, Oxford University Press.
Yang, Q. (2009) *Activity recognition: linking low-level sensors to high-level intelligence*, in proceedings of the international joint conference on artificial intelligence (IJCAI), pp. 20-25.
Young, S.W.H. (2014) *Improving library user experience with A/B testing: principles and process*, Journal of library user experience, 1(1).