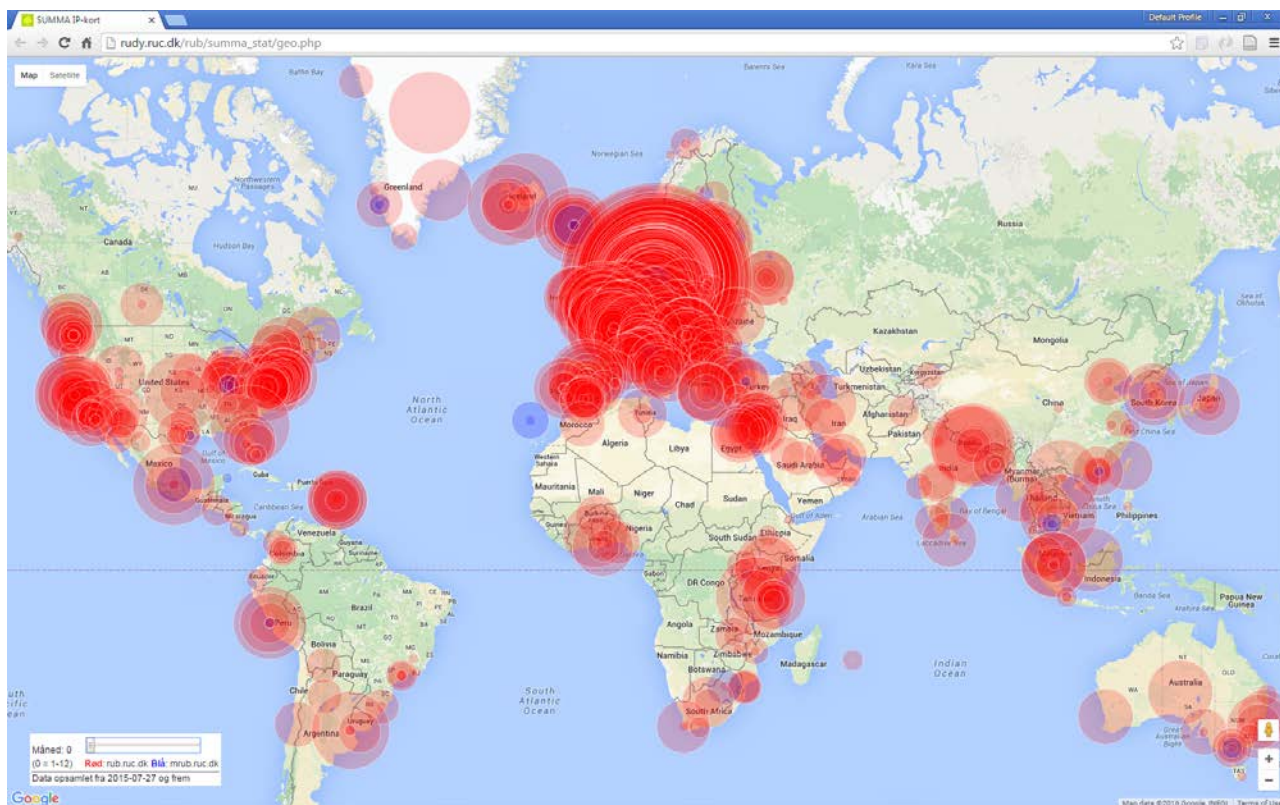# A DANISH LIBRARY IMPACT-REPORT

Results from Roskilde University Library

English summary

2016

Søren Møller and Peter Søndergaard, Roskilde University Library

Publication series from Roskilde University Library no. 63

A DANISH LIBRARY IMPACT-REPORT
English summary

2016

Summary from a DEFF project

Søren Møller, Roskilde University Library[1], Peter Søndergaard, Roskilde University Library[2]

This report is an extract of the results of a Danish library impact project. The main report is published in the *Publication series from Roskilde University Library* as no. 62

The cover picture is a result of the generation of spatial data from the use of Roskilde University Library's *Quick Search* (both the website and the web app). The data is generated using automatic translation of the log files' IP addresses to obtain geographical coordinates. The larger the circle, the more traffic.

---

[1] moller@ruc.dk
[2] pso@ruc.dk

Introduction

The library's impact studies relate to a new type of library studies, which was introduced around 2010 in the UK, USA, Australia and, unnoticed by most people, also in Iceland with a single study. At about the same time, a working group in ISO - ISO/TC 46/SC 8/WG10 under guidance of Roswitha Poll, undertook to draw up a standard in the area, and it was partly due to this work that the initiative came to Denmark. The library's impact studies or *library analytics,* which is the most common term for the phenomenon, refers in principle to all user-related studies, both quantitative and qualitative, whose purpose is to detect an effect of the library's services for/on the users who use them. This description is also contained in ISO standard 16439. Nevertheless, and unlike the standard method of placing all types of user surveys on an equal footing, the library's impact studies were from the outset a type of investigation that departed from the tradition of measuring *expressed meaning ("*what the library means to me is...") to measuring *demonstrated meaning* for the users. In particular, the intention was to measure the expressed meaning of students who use academic and research libraries. The new types of studies were often characterized by "going behind the backs" of the students and using the quantitative data that were stored on log-files and other electronic platforms in the educational institutions, as tracks of the users' library and study behaviour.

Those who developed the idea for the library's impact studies had noted that libraries had to a greater or lesser extent, and more or less randomly, collected and stored data on what users do when they are in contact with the library and use its services. This often involved extremely large, unorganised volumes of data, with and without personalised information, and the impact-pioneers, who undoubtedly were inspired by commercial sectors' analyses of *big data,* had to systematically review their own data and explore what they indicated, and if they could be systematised. There was still a long way to go to reach the objective of libraries themselves, based on the usage data, reflecting on the impact of their services and developing an approach based on electronic data that could be used to adapt existing priorities in more customized directions.

Conclusion and perspective

The examples show that *1) over time, the users of Roskilde University Library display a significantly altered behaviour in the use of the physical library. and 2) the consistent feature is that the library's users are characterized by generally performing better in their studies than non-users.* Users who use the guidance service *Book a Librarian* get higher grades than students who do not use this possibility. It should be emphasized that there is no evidence of a causal relationship, i.e. it cannot be concluded that students obtain higher grades because they make use of this library service. It can be concluded, however, that students who use the library's services are characterized by achieving higher grades than students who do not use the library.

The examples also clearly show that library impact analyses are suitable as a partial basis for decisions in the libraries to support and encourage demonstrated behaviour changes in users, by designing the physical library within a creative library concept, and that libraries can justify giving greater priority to librarian-supported learning for the students, because the user-group thereby perform better in their studies.

It can also be concluded that if library managers wish to use library impact analyses as a tool for the design of their service portfolio, the libraries must necessarily define a data policy specifying the types of user data they collect and save, and what user data will be possible to collect and save. The impression obtained during the project is that libraries differ greatly in their handling of generated data, and that the area is characterized by randomness, because the data has been regarded as a by-product and not as a key source for determining the libraries' service policy.

Notwithstanding that the project topic of *library impact* may invite a view of the more geeky side of the libraries' activities, the project has found that there is a keen interest in this new type of library studies. This is certainly because the library employees want to know whether their efforts are successful, whether they make a difference and whether their workplace is performing particularly well. After many e-mails, meetings and presentations, the project found that many library employees would like their own library to be able to *measure* the effects of different services more accurately, or at least based on quantifiable variables. Many libraries have experience with carrying out small surveys based on qualitative data or opinion data, whereas there is little knowledge about how to implement the new types of impact studies.

The following section with examples of Danish impact studies will provide a fact box for each example, which constitutes the toolbox. The fact boxes contain technical specifications for how the data is gathered, and there may be ethical and administrative considerations linked to the data collection. Alongside the concrete demonstration that the academic libraries actually make a difference to those who use them, the tool boxes are the project's main contribution, disseminating knowledge about how it is often possible to employ simple methods to conduct your own impact study. The tool boxes are markedly positioned below the examples of impact measurements.

Library impact - Danish experiences[3]

---

[3] There are not many examples of Danish impact-like library studies before this DEFF project. It should be mentioned, however, that the *Think Tank Libraries of the Future* has released the publication: *Folkebibliotekernes samfundsøkonomiske værdi*, 2015. [The socio-economic value of public libraries] It is a review of methods to determine the user-value of library services.
http://fremtidensbiblioteker.dk/temaer/hvad-er-folkebibliotekernes-samfundsokonomiske-vaerdi/
The purpose of the report is to assess the role of public libraries in creating socio-economic value. A Swedish forerunner was the report *Bibliotekens förtjänst – En forskningsöversikt om den samhällsekonomiska nyttan av bibliotek* [The benefit of libraries - A research overview of the socio-economic benefits of libraries]. Swedish Library Association 2007.
http://www.biblioteksforeningen.org/material/bliotekens-fortjanst/

Data sets in library analytics

Since the services offered by libraries are not organised in such a way to readily allow library impact studies, but are based on other considerations, this means that it may be random which studies are even possible to implement. It has largely been considerations relating to convention, reconstruction and security that have determined the collection and storage of library data. It is therefore random whether it is possible to find meaningful data sets or correlations between parameters in the data set.

Examples of the types of electronic data stored in log-files at RUb

Electronic data (e-data) is logged and entered data that is stored in log files. Log files are preserved in whole or in part on servers, depending on the IT procedures and back-up applications. A combination of chance, procedures and safety considerations determine which e-data a library has saved on their and/or the operator logs. When a library decides to implement library impact studies based on e-data, and that such studies should become an integral part of the library's strategic preparedness, this means that the library must organize its e-data policy accordingly, and that it *collects, stores, organises* and *safeguards* its data according to the requirements of the analysis.

Overview of existing log-files in RUb before the impact study

Impact						
Logningsdata med interesse for Library Impact og brugeradfærd						
Hvad logges	**Beskrivelse**	**Persondata**	**Format**	**Server**	**Placering**	**RUb**
Kviksøg	Kviksøg søgninger, dato, tid, antal hits, da/en, mobile	Nej	SQL	rudy.ruc.dk	MySQL hans_2 summa_log	RUb
Magasinafhentning	ID, titel, type, dato, tid	En uge	SQL	rudy.ruc.dk	MySQL magasin magasin	RUb
Reservationer	Bøger med mere end tre reservationer	Nej	Plain text	rudy.ruc.dk	/var/www/html/res	RUb
Gammel dørtæller	Antal ind- og udgange	Nej	Excel	http://rub.dancount.net		RUb
Ny dørtæller	?					RUb
Dspace	Download af PDF fra RUCs Digitale Arkiv	IP	Apache	http://diggy.ruc.dk	/home/thtoft/logsfil	RUb
Dspace	Upload af PDF til RUCs Digitale Arkiv	IP, bruger		http://diggy.ruc.dk		RUb
OJS	Download af PDF fra RUbs tidsskriftsystem	IP	Apache	http://rossy.ruc.dk	/home/thtoft/logsfil	RUb
DDELibra	?					RUb
Tynde klienter	?					RUb
Print	?					RUb
Kopi	?					RUb
Scan	?					RUb
Lokalebooking	?					RUb
SSO	?	username?	Apache?	http://molly.ruc.dk		RUb
Bestil i bibliotek.dk	?					RUb
VPN	?					RUb
Reftracker	Spørgsmål					RUb
Refchat	Chat					RUb
MS Exchange	Benyttelse af lokaler					RUb
Indlån	?	Lånernr				RUb
Bogforslag	Forslag til indkøb af bøger	Lånernr				RUb
SUMMON	?		Google analytics			RUb
PDA	Brug af PDA: køb, lån, browse	username?				RUb

EXAMPLE 1: Number of visitors and duration of visit

Roskilde University Library has continuously measured and stored data about the number of people passing through the entrance door since 2002. There is only one entrance, so only one counter is required. The counter can distinguish between people entering and exiting and has a time resolution of 1 hour. The number of visitors and the times during the period from 7 November to 7 December has been used each year since 2002 to analyse changes in the users' behaviour. Thus, an overall decline was registered in the num-

ber of visitors up to 2009, after which there was a slight increase and a stagnation after 2012 (Figure 1: note that there were problems with the entrance counter in 2003 and 2004).
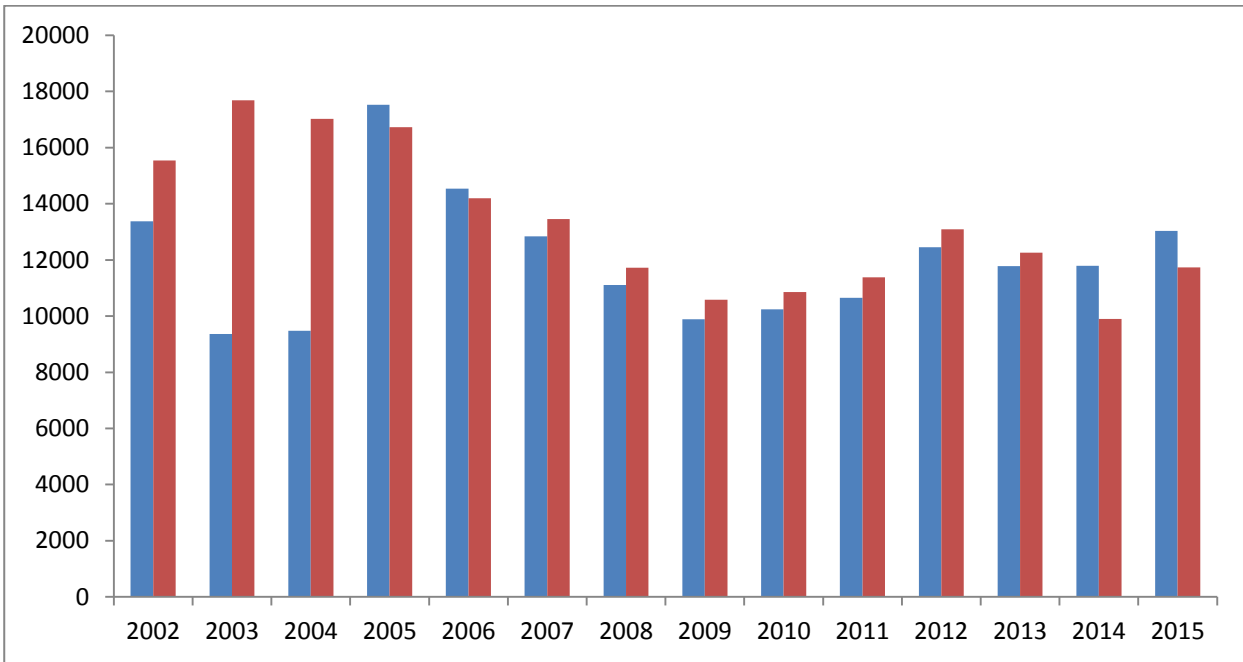

Figure 1. Visitors to RUb 7/11-7/12, 2002-2015. X-axis: years, Y-axis: number of visitors. Blue columns: entering, red columns: exiting. The differences in the red and blue columns within a year are signs of a non-critical lack of precision in the counting apparatus.

RUb's visitor data can also be used to estimate the *duration of visits* to the library, the retention time, by calculating the difference between the medians for the entrance and exit times (Figure 2).
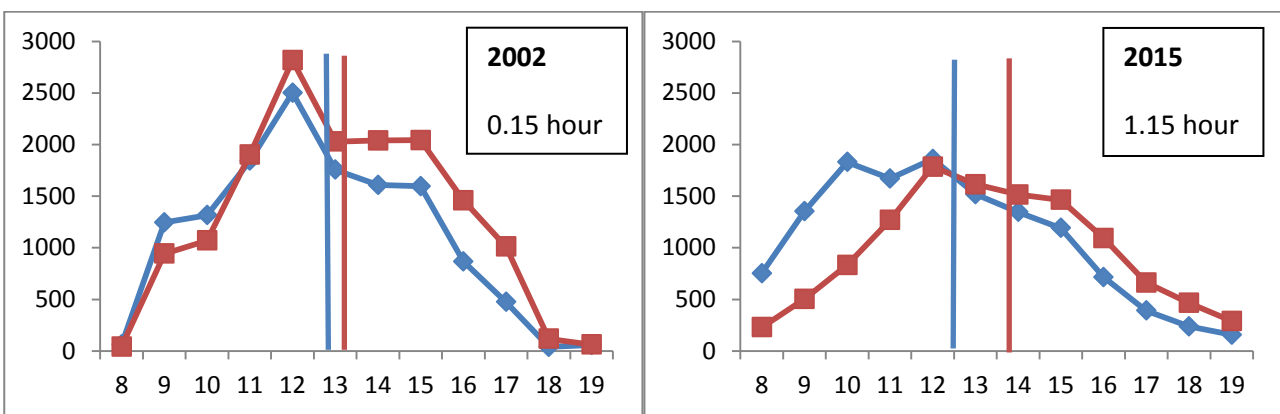

Figure 2. Changes in the duration of visits to RUb 7/11-7/12, 2002-2015. Distribution and medians for entrance (blue) and exit (red) times in 2002 and 2015.

There is clearly a marked increase in the duration of visits to RUb (from 0.15 hours in 2001, to 1.15 hours in 2015). Figure 3 shows the duration of visits for each year throughout the period. The red "tips" for 2013, 2014 and 2015 show the duration of visits, if we also include the unattended hours, which were introduced in 2013.
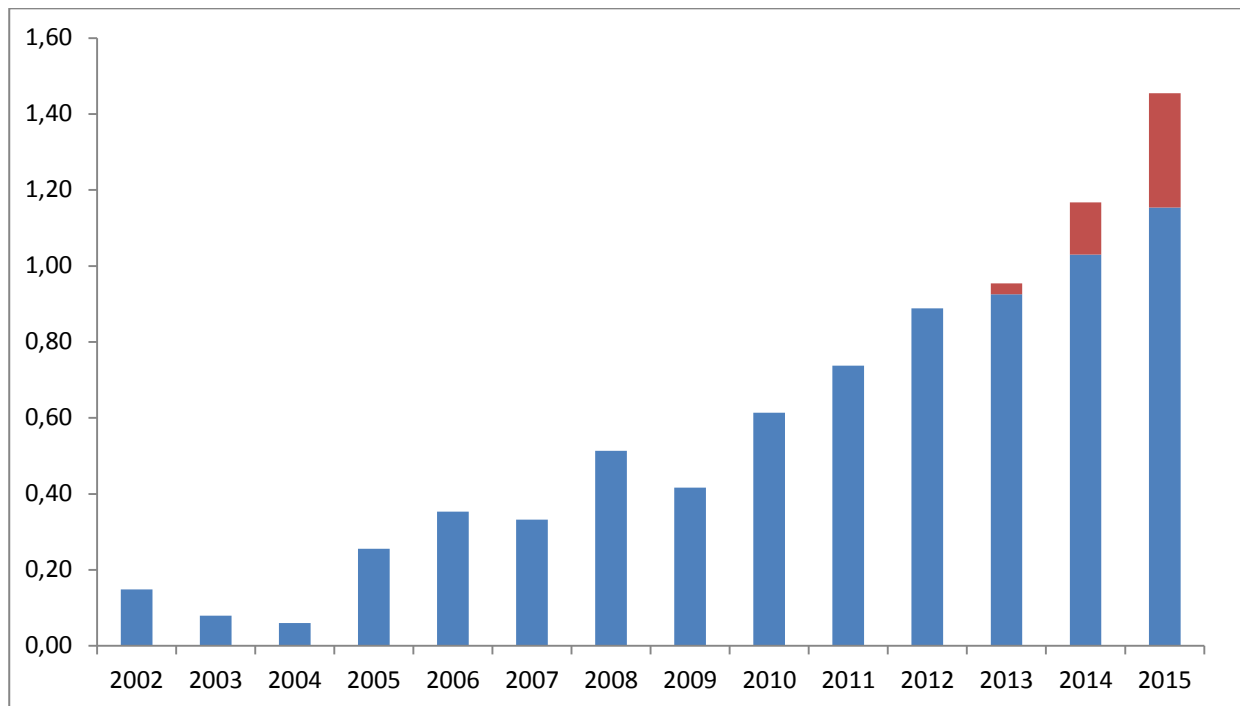

Figure 3. The blue columns show the duration of visits calculated as the difference between the medians for the visitors' entrance times and exit times. X-axis: years, Y-axis: average duration of visit. The red part of the columns indicates the proportion of the visiting time that was outside the periods in which the library is staffed

Tool-box: Visit time in libraries

Time spent in libraries when physically visited. Visit times may be estimated through interviews, counted by Infrared (IR) or similar detectors, or by detecting the visitors' WiFi-devices.

Tool-box: Library visit time counted by IR detectors:

The library entrance is equipped with IR-sensors that are able to distinguish and detect incoming and outgoing persons. In the market for counters, some devices have facilities that calculate an estimate of the average time spent in the library by visitors (or customers), i.e. the *retention time*. Other counters lack such a facility, but they collect data with a specified time resolution down to one hour for ingoing and outgoing traffic. In that case, it is possible to estimate the average visitor's retention time. For a given period (e.g. 30 days), calculate the medians (or mean if data is normally distributed) for ingoing and outgoing times. Medians for grouped data are calculated from the formula:

$$Median = I + \left(\frac{\frac{n}{2} - cf}{f}\right) * h$$

l = lower limit of median class

n = number of visitors for the whole period

cf = cumulative frequency of class prior to median class

f = frequency of median class

h = class size

The difference in the two medians (or averages) for the ingoing and outgoing times constitutes an indicator (proxy) of how long users are staying in the library, the *retention time*.

Retention time can also be calculated from data collected from WiFi equipment. Data is collected with a Raspberry Pi Linux computer and stored in a MySQL database. Web-interface is used for display and analysis. RUb has data from 07.11.2015.
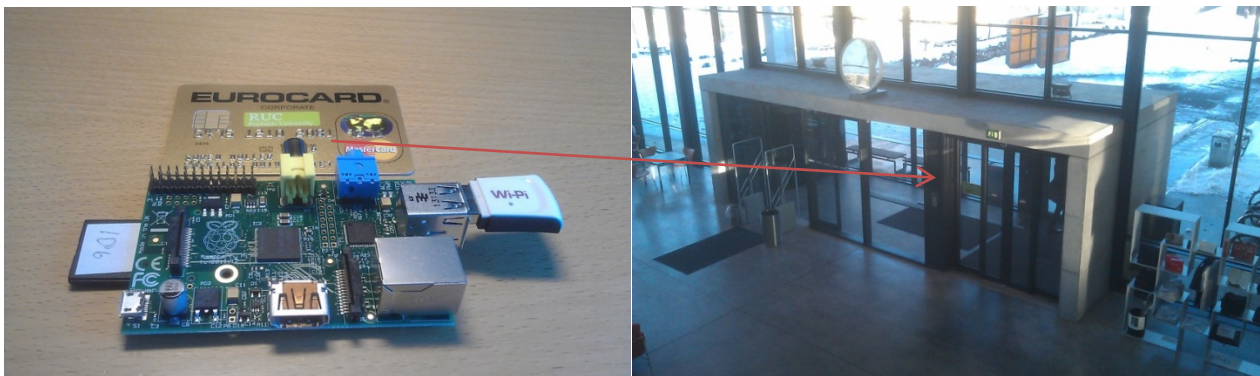

Figure 4. Raspberry Pi Linux computer and its location at RUb's entrance area

Tool-box: Library visit time detected via WiFi-devices:

All devices with activated WiFi network interface (mobile phones, tablets and laptops) continuously transmit signals which provide information on the unit's unique Media Access Control address (MAC address). These signals can be captured and MAC addresses can be extracted and stored (in encrypted form for the sake of privacy) in a database.

We have tentatively used a Raspberry Pi 2 Model B Linux computer (cost about €110), equipped with a WiFi adapter and the open source programs airmon-ng and tshark to capture and extract MAC addresses detected at the library entrance. MAC addresses are encrypted in a PHP script, which also takes care of transmission of the encrypted data to a database server, where they are stored in a MySQL database along with a time-stamp and an indication of signal strength. In this experiment, we have followed individual devices (and thus their users) retention time in the library, defined as the first and last time-stamp of registration in the course of a day. The retention time can then be calculated as the difference between the average of the first and last time-stamp of registration for all MAC addresses measured over one or more days.

For the processing and display of data, we have developed a web interface for automatic calculation and graphical presentation. The web interface is programmed in PHP with jpGraph for generating graphs.

The system offers a wealth of opportunities for further analysis. For instance, one could look at the distribution of how many times individual users visit the library during a given period. If WiFi detectors are installed in all the rooms of the library, it would be possible to track users' movement pattern. In larger rooms with multiple devices installed, the movement patterns may be followed in great detail via triangulation based on signal strength.
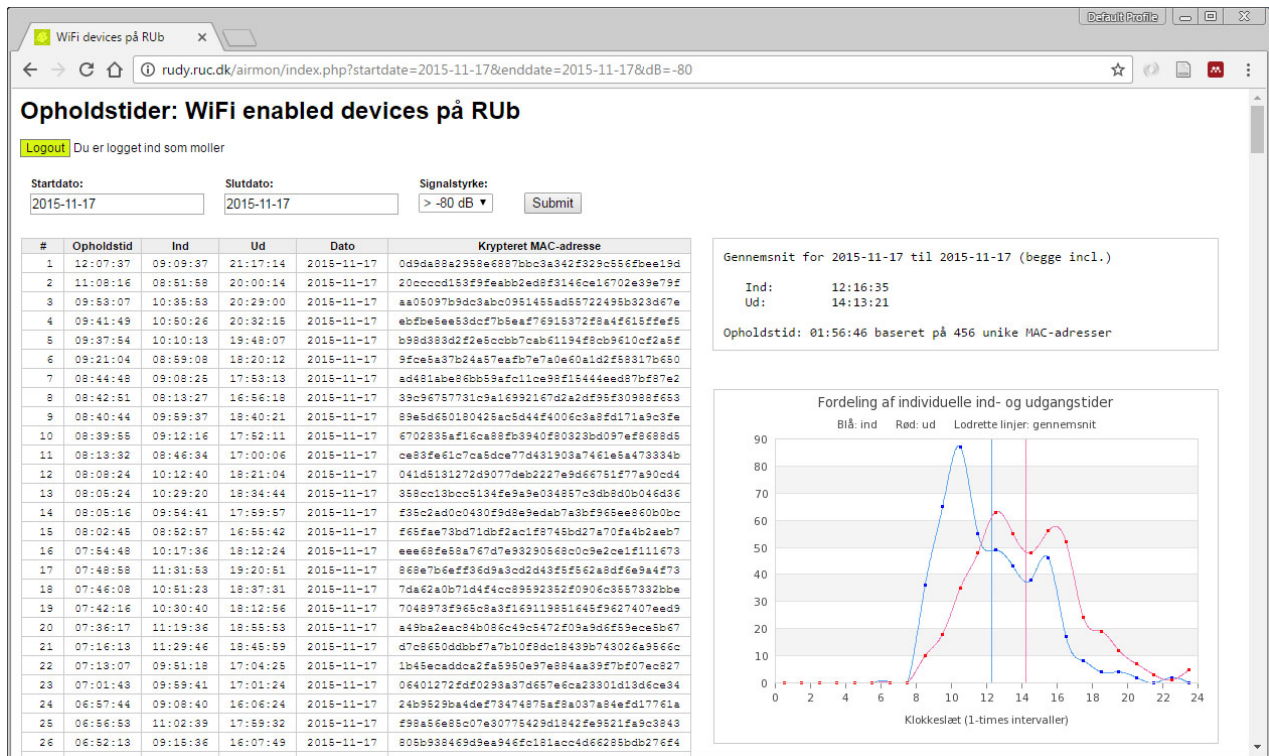

Figure 5. Web interface for display and analysis of Wi-Fi data

Measurements of visit duration based on WiFi data produces a result of 1.87 hours for the same period as shown in Figure 2, right side (7 November to 7 December 2015), with data from a mechanical door counter, and showed an average visit duration of 1.15 hours. It is predictable that results based on WiFi data will exhibit slightly longer visit duration, because people who visit the library for a short period are unlikely to have turned on laptops or tablets, while those who spend longer periods at the library are likely to activate more WiFi devices. However, Figure 6 below shows that the data is still largely characterized by WiFi devices that are only registered for a short time. The method also allows measurement of the distribution of MAC addresses (devices) over time and the distribution over days, and unsurprisingly we see the largest number of devices in the early afternoon (13:00-14:00). Figure 10 shows the distribution over days during the measurement period, where weekends clearly show low visitor numbers, completely in accordance with data from the door counter.
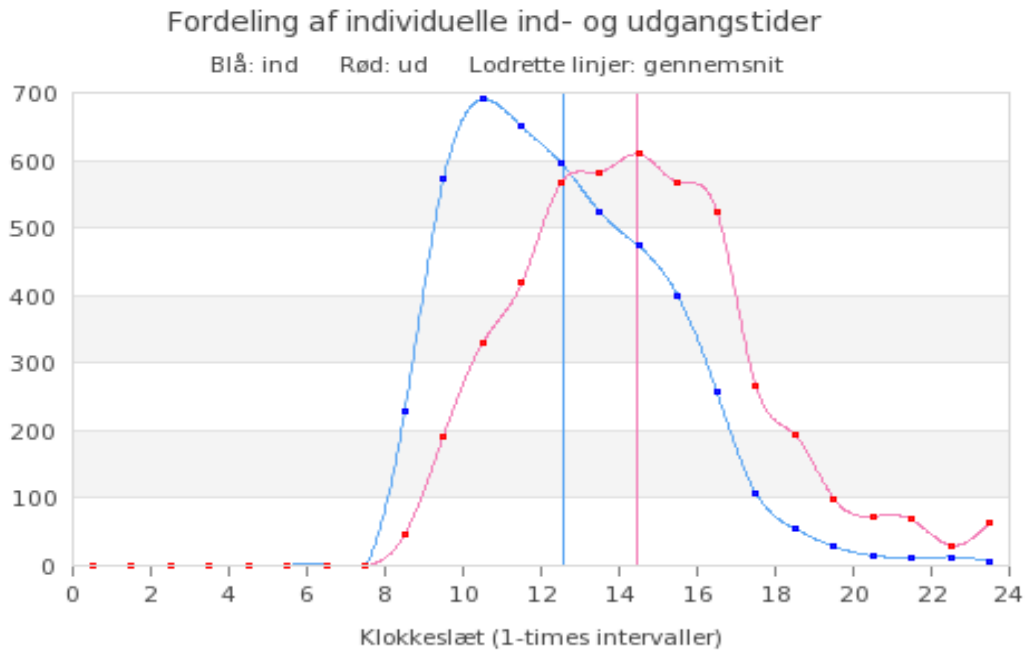
9

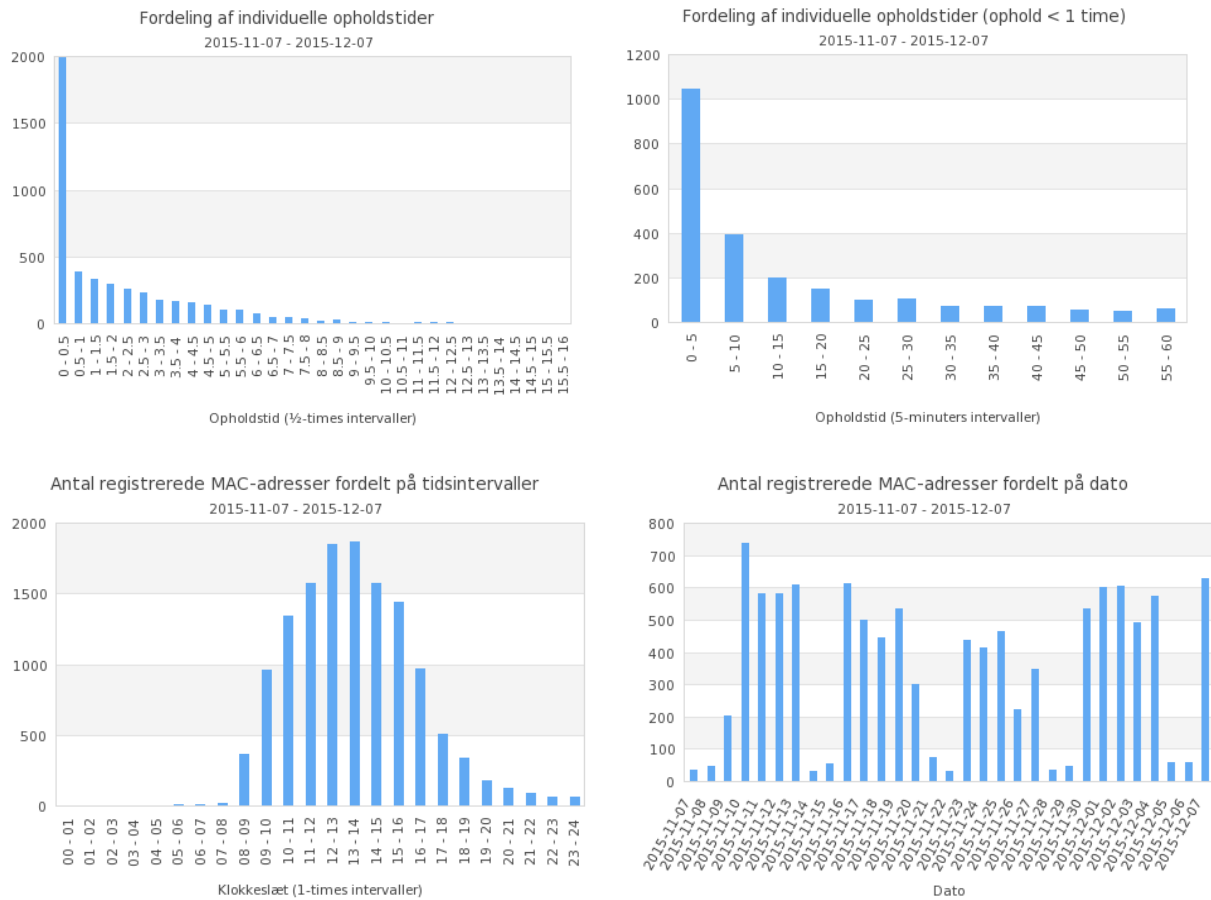**Figure 6. Analysis of Wi-Fi data**

Figure 7-10. Analysis of Wi-Fi data

RUb's visitor data is characterized by being long series of time data, without any individual personal identification of the visitors, since we disregard the Wi-Fi addresses of the visitor's IT devices. If we wish to collect data, where personal identification is possible and the visitors can be identified, this would require visitors to check in or otherwise be identified when they enter and exit. This is not the case at RUb or at any other public library in Denmark, with the exception of periods during the day when the libraries are not staffed. Therefore, in the case *RUb visits*, it is not possible to conduct any kind of impact measurement that can be linked to other individualized parameters. Nevertheless, this type of measurement can have impact consequences.

The history behind the visit and retention data for RUb is that a decrease in the number of visitors was detected after 2002. This was shown by the visit counter and it gave rise to some dismay and anxiety with regard to the physical library for several years. But while the number of visitors declined, the staff could see that during some periods of the year, areas of the library were full of visitors, and a suspicion gradually developed that the counter was defective. It was decided at one point, possibly inspired by supermarket advertising about how fast it was possible to do your shopping, to supplement the figures for visitor numbers with information on the duration of visits to the library building, and it turned out that the actual duration of visits had increased dramatically during the period in question, and that this was the explanation for the apparent paradox.

In what way can we speak of an impact in relation to visitor and retention data? RUb *interpreted its visitor data* as evidence that the public thrived in the library premises, because they could concentrate and be inspired by spending time in a simple, quiet architecture. Based on this interpretation, it was a good idea to consider how the physical library could be adapted to make it even more attractive for study purposes. This interpretation has had and continues to have major influence on how design considerations take place at RUb. The correctness of the interpretation is confirmed by the recurring study environmental surveys. It applies to all measurements, where the data is isolated, that attempts to interpret their significance and consequences require argued contexts and that these possible contexts are constantly challenged with new observations. Generally, it must be said that, if we wish to promote new usage patterns, a dramatic extension of the duration of visits in the physical library, like we have experienced in RUb, calls for a significantly changed design intended to inspire and facilitate spending more time in the library. This is what RUb has tried to achieve.

Visitor and retention data can also be used when information and guidance preparedness needs to be configured on a daily and semester basis. Visitor and retention data can probably also be included in the planning work at libraries that plan based on an annual cycle.

EXAMPLE 2: Search behaviour

The use of Roskilde University Library's *Quick Search* is recorded in a log-file, which logs search strings with timestamps, the number of hits and the user's IP address is logged (for the record, no information is logged that is not also logged by the web server). This type of logging makes it possible to examine user behaviour in detail by using the SQL search language. Figure 11 shows the number of searches distributed over months since 1 January, 2011. The number of searches exhibits a general increase from year to year, but also shows clear semester fluctuations, where the autumn semester is the busiest.
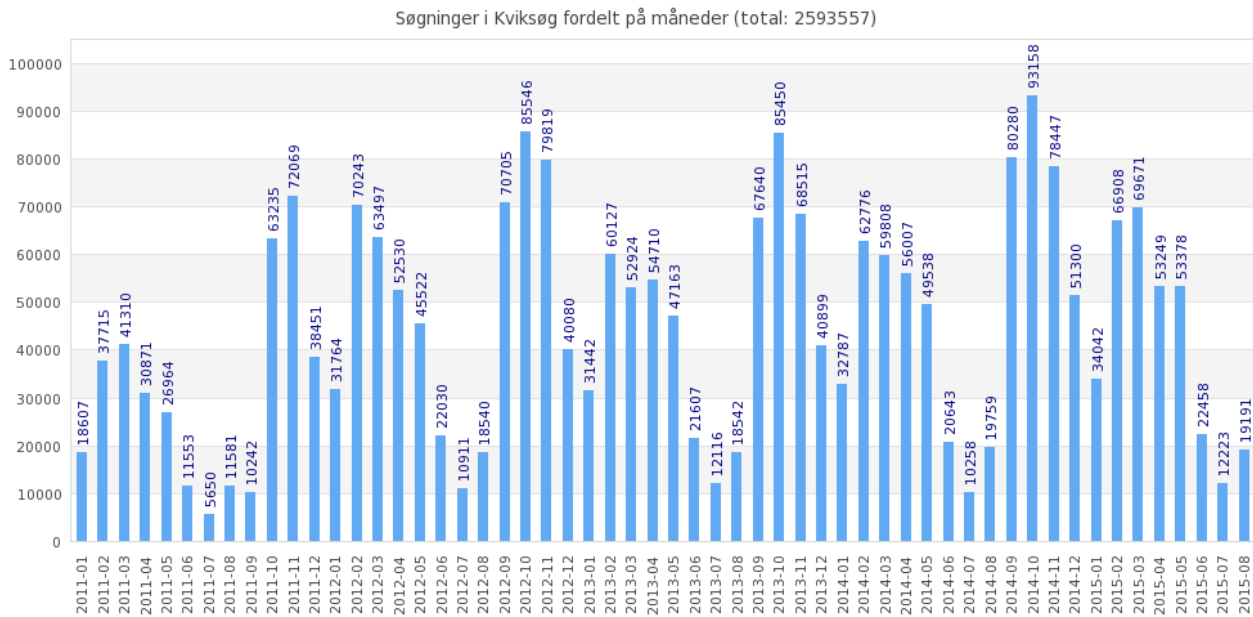
Figure 11. Periodic rhythm in Quick Searches at RUb 2011-15

Like the visitor log, the Quick Search log contains no data that can identify an individual person, and the impact is therefore based on interpretation. An "annual cycle" can clearly be observed in the data accrual, but careful consideration must be given to whether it can be used in the work planning for purposes other than to choose the least busy times for any IT conversion days. However, the search strings entered by the users can be used to analyse the users' search behaviour in a variety of details. For example, Table 1 shows the 20 most common searches in 2015 until the end of August ("logbas" is a search command that is added when the user uses search filters).

It appears that the most frequent searches are a mix of word searches for titles, authors and concepts. Unsurprisingly, there is no systematic subject searches among the most common. Library systematics is only a system for experts, library staff, and as a tool for shelf layout. No one else wants to use it or learn it, despite many years of persuasion. As might be expected, the main trends in the searches are that they closely resemble the questions that students ask their project supervisors, but surprises can be found even among the top ranked searches, which could suggest a need for a more accurate introduction to the library and for a reorganisation of the website. This applies, for example, to searches on topics such as "infomedia", "Rudar", "video", "speciale". Although these searches produce results in the Quick Search, it would have been more appropriate to search in a different fashion or to have had the opportunity to be referred elsewhere on the website, because it must be presumed that those who searched for "infomedia" wish to search for particular newspaper articles, and those who search for "Rudar" wish to search for student reports etc. in RUC's digital archive.

It is not impossible to imagine that the Quick Search log could be supplemented with personalised data. It would also not be a radical shift if users were asked to log in before they performed a Quick Search, given that the same users must log on when they request access to external licenses outside the campus. They

would hardly notice that the entrance barrier was moved slightly forward if they were required to log on earlier to perform Quick Searches. However, this would require a decision that had not been taken when this project was launched, but it would be a necessity if the use of e-licenses is to be measured correctly and personalised, regardless of whether the searches are performed from the campus or elsewhere.

Antal	Søgestreng
625	infomedia
272	kvalitative metoder
254	kvale
253	interview
241	videnskabsteori
230	bourdieu
223	Rudar
210	speciale
200	(kvalitative metoder) logbas:"2"
198	video
180	foucault
162	(speciale) logbas:"4"
161	diskursanalyse
157	sklerodermi
140	fairclough
136	Goffman
136	Kommunikation
131	deltagende observation
129	projektledelse
127	(Kommunikation) logbas:"4"

Table 1. The 20 most common searches in Roskilde University Library's "Quick Search" in 2015.

Tool-box: Facts about search behaviour

RUb's "Quick Search" is a locally developed web interface for the library's "discovery service" (catalogue and other databases), which is based on SummaRise from the State and University Library. This allows all searches to be logged directly in a local SQL database, where the search strings are stored together with a time-stamp (date and time), the users' IP address, interface language (Danish/English), interface type (workstation/mobile) and the user's geographic coordinates (see Example 3).

Analyses of the users' search behaviour can be extracted from the SQL database, either by manual SQL queries or through a web interface (programmed in php), which can also display the analysis in graphic form (Figure 11) using JpGraph (addition to php). Since the data is stored in the database instantly, the analyses will always be updated, both for direct SQL queries and in the web interface.

It is conceivable that in future, spatial data (Example 3) could be connected to search behaviour to see whether, for example, there are differences between on-campus and off-campus searches, or variations in searches from different interfaces (language and device type). The latter could provide an indication of how RUb's future responsive web design should be configured.

EXAMPLE 3: Spatial data

It is possible, with some precision, to "translate" users' IP addresses (from the Quick Search log) to geo-graphical coordinates. Very precise spatial data can be obtained from mobile devices, with the user's permission. Since the end of July 2015, RUb has logged this data in connection with the use of the library's "Quick Search" function, and the data can be plotted on Google Maps to create an overview of where the users are located when they perform a search. The map below (world, Copenhagen metropolitan area, RUC's campus) illustrates the number of searches performed in June 2016 from XXX unique geographical positions. The diameters of the circles are logarithmically proportional to the number of searches from the location in question. Red circles are based on IP numbers (searches on the library's full website, rub.ruc.dk), and blue circles are based on mobile devices' geo-localization (searches on the library's web-app, mrub.ruc.dk). It can be seen that "Quick Search" has users on every continent, but a very large proportion of the searches are conducted in the Copenhagen area and on RUC's campus. Many of the searches are from mobile devices located in the Trekroner area, on the campus and in the library building (Building 26).


Tool-box: Facts about spatial data

IP addresses can be translated to geographical positions with a precision down to city-level. There are several on-line services (APIs) available for this purpose, but a database is also freely available on maxmind.com (GeoLiteCity.dat) with geographical information based on IP addresses. We have used the GeoLiteCity database for automatic translation of the log-files' IP addresses to geographic coordinates via php-scripts. Mobile web-apps allow a high degree of precision and (with the user's permission) provide information about the user's location and can then log this information. We have used these methods to collect spatial data from the use of Roskilde University Library's Quick Search function (both the website and the web-app) since 27 July 2015. These data are then plotted using Google Maps JavaScript API (https://developers.google.com/maps/documentation/javascript/).
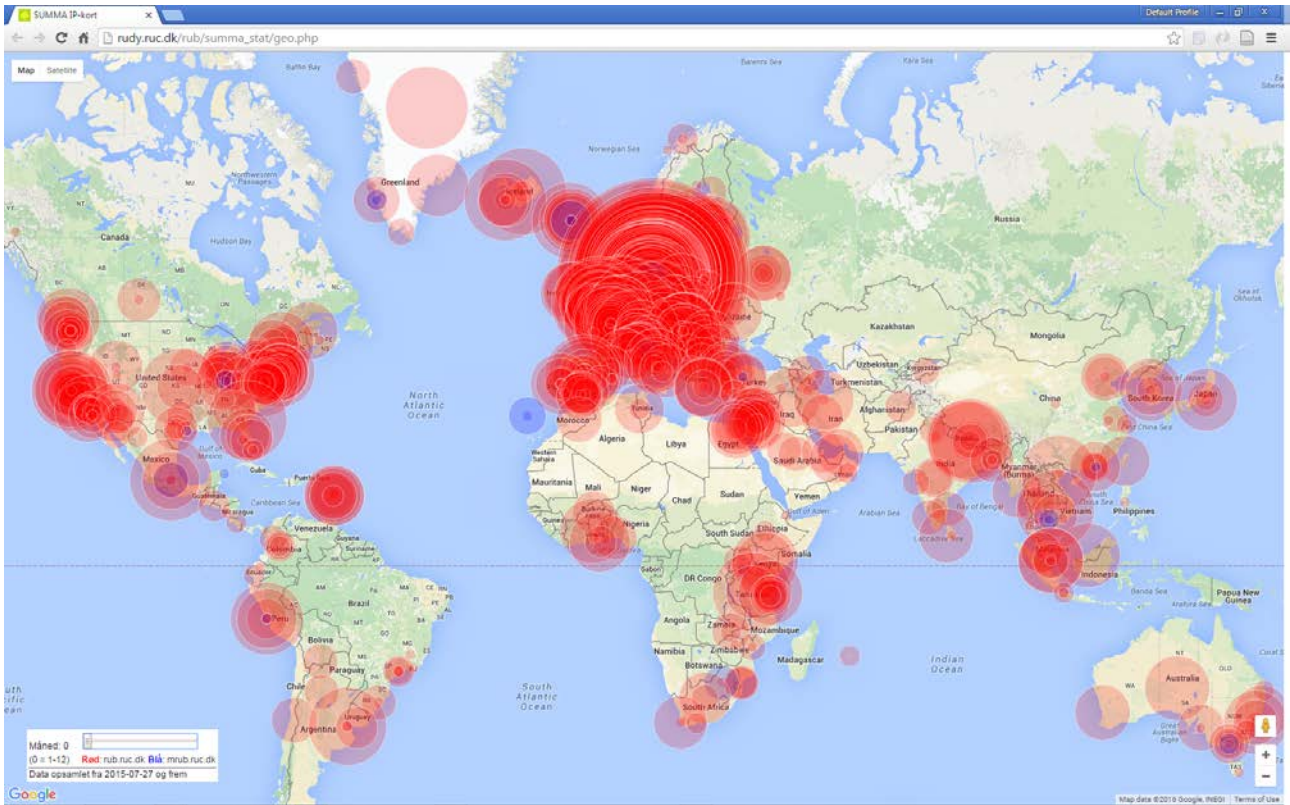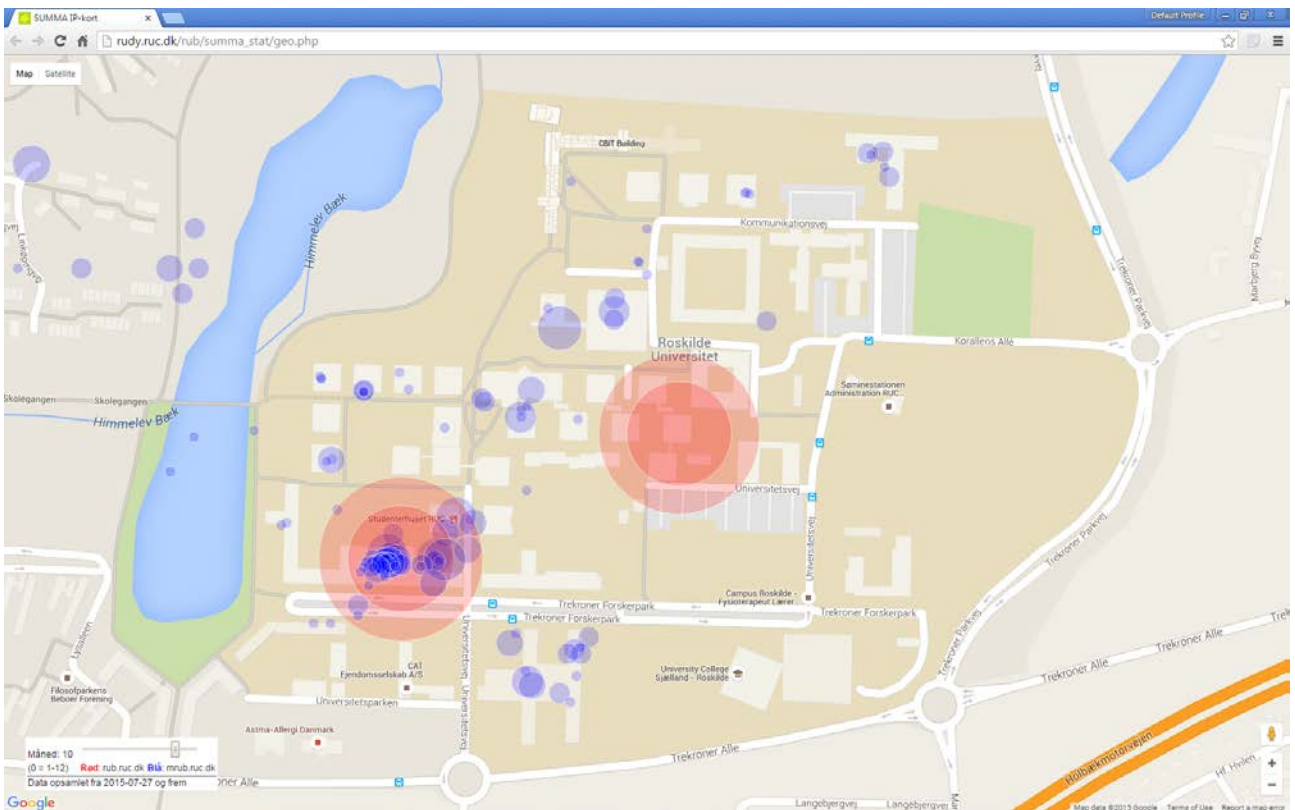

Map 1 Plotted spatial data for searches in RUb's Quick Search, July 2015 to June 2016 - World
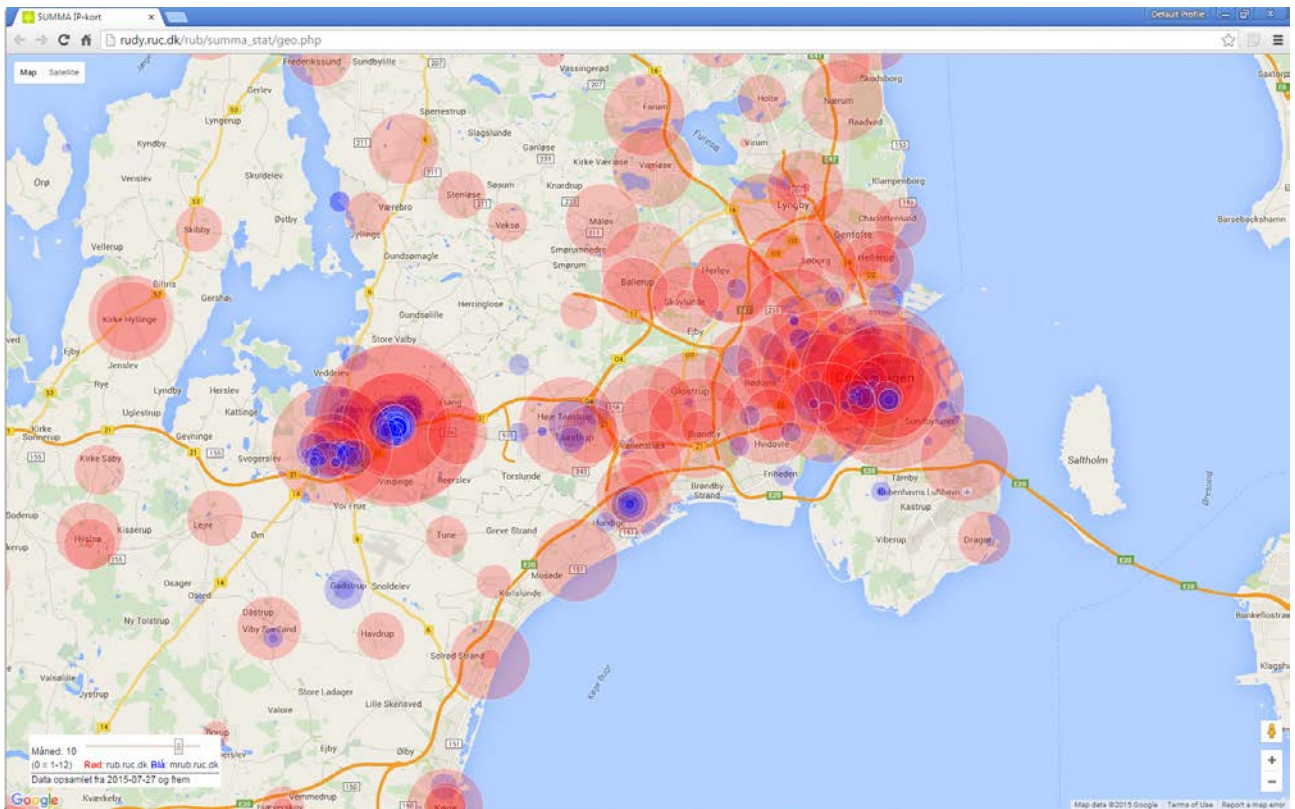
Map 2 Plotted spatial data for searches in RUb's Quick Search, July 2015 to June 2016 - Copenhagen metropolitan area

Map 3 Plotted spatial data for searches in RUb's Quick Search, July 2015 to June 2016 - RUC's campus

What importance can spatial data have for arguments about library impact? It cannot have a great deal of direct influence, but the spatial data is surprising and gives rise to reflection on how to rethink the service concept and introduce innovative ideas. For example, it is surprising that there are so relatively many users in Africa and South America using RUb's Quick Search. The reason may be that there are students and researchers from RUC's "development studies" who are present in these parts of the world, but we do not know yet. Spatial data can thus help us to differentiate and learn about the user segments, who voluntarily or involuntarily only use Quick Search and e-documents. If we know who they are, we may also be able to develop a library service that is more focused.

There is no available connection between IP/MAC addresses and personal ID, so we cannot link spatial data with personal data. But we do have the option to initiate contact with the remote user of Quick Search, for example by presenting the user with a pop-up with a text like: *We can see that you are in (country/region/continent), and yet you are using RUb's website. We find this rather interesting. Would you be willing to answer a brief questionnaire about your use of rub.ruc.dk? If so, click here.* The library knows very little about the user segment that uses remote access to the site and resources. Naturally this situation should be rectified, and we plan to address the issue. If these users are RUC students in internships or work placements, or RUC researchers on e.g. research assignments, we could consider asking them whether the library can offer them a special service, e.g. offer them special assistance related to the fact that they are not physically in the vicinity of the library's self-service functions.

EXAMPLE 4: Exam results and Book a Librarian

One of the crucial problems for this type of study is the fact that data, which allows the library service to be linked with personal data, is rarely available in an organised form, and that data collection requires planning and time. We are fortunate in this DEFF impact project that RUb has archived data over many years and, when it is combined with other data, it can shed light on whether a very popular service for students and staff actually has any measurable positive effect. The service is called *Book a Librarian* (B1B). From the point of view of its content and categorisation, it lies within a grey area between formalised library search instruction and actual project supervision. The service typically involves a group of students who request assistance to search for information for their bachelor project, or a new PhD student who needs help to find literature relating to a defined theme. At RUb, the B1B scheme entails that project groups, or thesis students and researchers, describe their requirements on a special on-line form, where they specify the problem definition, the project field, theory etc. This request is filed in the web platform *Reftracker* and can be accessed by the staff responsible for the B1B service. Then the librarian or information specialist arranges a meeting and prepares for the meeting with the students, which will:

- Explore the search options available and demonstrate how they can be used optimally. For example, RUb's Quick search and RUb's resources, bibliotek.dk, subject-relevant reference works, statistics etc.
- Help with the choice of search strategy and search tools.
- Select relevant databases and demonstrate their possibilities and limitations.

- Find other relevant sources and offer advice for evaluating internet sources.
- Provide advice on reference management, bibliographies and tips for documentation of the search process

As these inquiries have been stored in a data archive, it has been possible to manually transfer the cleansed data to an excel worksheet, where students and project information have been supplemented with data about the study progression of students who used B1B and those who did not. All the data was then gathered on a special, secure server, and the data was analysed for possible correlations.

The database contains a great deal of other data, in addition to the B1B data, as it includes all the inquiries that have arrived at RUb via the *Reftracker platform* for B1B, chat or e-mail. The selected period covers data from September 2014 until the end of December 2015, i.e. a total of three semesters or almost 1,450 inquiries. The e-mail inquiries that do not request a B1B session, mainly concern requests for assistance to organise practical issues concerning physical loans, and e-mail requests for the loan/purchase of specific titles.

For the preliminary analysis of B1B data, it was decided to focus on requests for assistance from bachelor students. The table illustrates the listed RUC-subjects for bachelor study programmes, while BP1, BP2 and BP3 indicates a bachelor studies project, which comes at the end of each semester for the first 1½ years of study. The students receive individual grades. The first two projects, BP1 and BP2, are graded through an internal assessment, while the grading of BP3 occurs with the participation of an external examiner. The digits in the boxes represent the study components, as they appear in STADS, the studies administration system, from which information about grades can be retrieved under certain conditions[4].

Exams/Study programme	HUM	HUMTEK	SAM	SIB	NAT
BP1 winter 2014/15	24810	24788	24859	24877	24763
BP2 summer 2015	24812	24791	24862	24880	
BP3 winter 2014/15	24814	24809	24869	24891	24767

Table 2. Numerical data for exams at RUC's bachelor programmes' 1st, 2nd and 3rd semester 2014-15

The preliminary analysis shows that there is a positive and statistically significant correlation between the average grades for the bachelor students used the B1B service, compared to students who did not use the service.

Effect/Impact of B1B on students' project grades at BP1, BP2, BP3 2014-15

gs = average; N=number	+B1B gs (N)	-B1B gs (N)	Difference	p (t-test)

---

[4] The condition under which the DEFF project can obtain access to STADS has been agreed with RUC's Administration. In general, access authorisation can be granted for research purposes, if the data is rendered anonymous when publishing, and such that the data is stored internally on a secured server.

+B1B/-B1B	8.95 (1509)	8.72 (2146)	+0.24	0.008413 (* Significant (p < 0.05))

Table 3 Average grades for RUC students who used the Book a Librarian service in 2014-15, compared to exam results for students who did not used the service during the same period

---

Tool-box: Fact box for investigation of the efficacy of library training (B1B)

In order to perform this type of comparative statistical studies of the efficacy of teaching and guidance, measured in terms of study progression/grades, the following are required:

A. Two or more defined and randomised populations[5]

- A full set of user data from a time-restricted period for those individuals who have received a service: teaching/guidance/B1B
- A full set of data from the same restricted period for all individuals in the entire potential circle of users, who could have received the service

B. The individuals in the population must be ID-identifiable and anonymized if necessary (e.g. social security number, student number, library number, study programme) [6]

C. Data from result/progression runs from STADS, distributed according to ID etc. (grades, study programme, gender, marital status, study progression, age, etc.).

D. Database for summarized and anonymized statistical calculations.

E. This database can later, and under certain circumstances, offer user access, user interface prototype[7], where students can see and consult the information that the educational institution has stored. (Studies in the UK (JISC dashboards access) and Australia (Cube)). Data in STADS is transferred either from various digital application portals or is entered manually. Data in STADS about the students consists of:

- Social security data/number (gender and age data)
- address - updated - plus previous addresses; e-mail (not used in the DEF project)
- nationality (not used in the DEFF project)
- exam grades (all?), approved courses
- leave of absence, exchange (not used in the DEFF project)

In STADS Online Services, the students themselves register for their exams (Students register for courses/exams via STADS Online Services). STADS does not contain data about the student's marital status, birthplace, financial circumstances, including any responsibilities for dependants. The student's enrolment in STADS is terminated when they leave the university.

---

[5] Regarding the entering of data in a Jisc LAMP project, refer to several contributions on
http://jisclamp.mimas.ac.uk/

[6] Lee Baylis: Unique Identifiers Which Don't Identify Anyone! (2013)
http://jisclamp.mimas.ac.uk/2013/06/28/unique-identifiers-which-dont-identify-anyone/

[7] Mike Jones: A Library Analytics and Metrics Service? Moving into the next phase of work (2014).
http://jisclamp.mimas.ac.uk/2014/08/22/a-library-analytics-and-metrics-service-moving-into-the-next-phase-of-work/

What can the survey results be used for? When, as in this case, a result is found to show that the B1B ser-vice helps students achieve a more satisfactory study progression, it is natural to report the results to those involved, in this case the relevant librarians and bachelor students, and especially their teachers. Next, the results must be discussed. Is the observed correlation satisfactory, or is the difference too small? Can the difference between the two groups be increased by improving the B1B-training? Can the booked librarians develop the content of the B1B training and increase the benefits that the students receive from the train-ing, in terms of both better understanding and better grades. However, they are aware that they are mak-ing a difference, and nothing could be more motivating than trying to make the differences more signifi-cant, for example by refining the content of the B1B training. This transforms the discourse from being a study to locate the library's impact, to being a targeted and evidence-based effort to develop a pedagogi-cal-didactic learning platform for more effective information retrieval training.