Accessing National Bibliographic Data in Visual Dialog with Biographic Data

Ylva Sommerland
National Bibliography Division, National Library of Sweden, Stockholm, Sweden
ylva.sommerland@kb.se

Abstract:

Strategy development and technical solutions for exploiting the full potential of national bibliographic data are central to ongoing efforts in efficiently meeting goals at national libraries. This paper present methods and models for evaluating the usefulness of national bibliographic data that rely on technical presuppositions, focusing on analyses of how the data's visual representation and narrative content may become more understandable to users and ultimately contribute to furthering knowledge. These methods were developed with the aim of enhancing the accessibility of national bibliographic data. The paper addresses the importance of transparency and documentation of criteria for unique data within a body of national bibliographic data. Qualitative goals for data in relation to user needs are demonstrated by a set of visualization experiments extracted from data derived from the Swedish national bibliography.

The National Library of Sweden began to develop methods for evaluating the usefulness of national bibliographic data by way of visualizations in 2015. In this paper it is discussed how the quality and usefulness of national bibliographic data may provide valuable insight into potential development and improvement focus areas, and in doing so, uncover new knowledge. The project was developed in three steps: 1) publishing a statistical report 2) visualization of findings 3) a fact-sharing seminar.

A conscientious renewal of the annual statistics report sought to convey the complexity and enormous span of findings, and potential knowledge gain possible to exploit from the data. What stories would surface from the national bibliographic data about Swedish publishing trends? How could methods be developed for users to interact with the data and create their own relevant data sets? Broader findings would not be immediately perceivable after only one year, so the goal was to limit the scope to a single theme, and plan for a recurrent, annual event to showcase findings over time. The theme chosen in 2015 focused on "life stories", and is based on a data set limited by biographical coding.

Keywords: national bibliographic data, book statistics, data visualizations, information visualizations.

1 A NATIONAL BIBLIOGRAPHIC STATISTICAL FLOW

*The lack of reliable-or even broadly realistic- data and analysis on
the world's book culture and publishing markets has been deplored time
and again. Yet, the gap has not been bridged.[1]*

The National Library of Sweden began to develop methods for evaluating the usefulness of national bibliographic data by way of visualizations in 2015. In this paper it is discussed how the quality and usefulness of national bibliographic data may provide valuable insight into potential development and improvement focus areas, and in doing so, uncover new knowledge. The project was developed in three steps: 1) publishing a statistical report 2) visualization of findings 3) a fact-sharing seminar. There were two main reasons why this project was initiated. Firstly, the methods of presenting the annual statistics report from the Swedish national bibliography had not been thoroughly evaluated for several years. A need to update the statistics for books registered in the Swedish national bibliography had clearly arisen during these years. Researchers and other stakeholders had recurrently called out for complete and reliable statistics on the Swedish publishing output. It has been an ongoing discussion and critique in Sweden for decades that there exists no institution that offer fully satisfying statistics on the publishing output, and as stated in the introductory quotation above, this is an international problem. It was important in the process of starting to develop methods and models for evaluating the usefulness of national bibliographic data to focus on analyses of how the data's visual representation and narrative content may become more understandable to users and ultimately contribute to furthering knowledge. Since 2011 the Swedish national bibliography and authority data were released with an open license. An important goal of the project was therefore to find out how methods could be developed for users to interact with the data and create their own relevant data sets.

Secondly, a plan in line with this was to make the data and the knowledge about the publishing output visible through a yearly event – where authors, researchers and the National Library together in a public fact-sharing seminar would open the door in to the stories of the national bibliography. This part of the project was inspired by *the Swedish Institute for Children's Books* and an event they call "Book tasting" – an internationally unique event arranged annually with the purpose of presenting national trends and analysis of the total children's and young adult book publishing in Sweden.[2] The National Library of Sweden likewise has access to a unique overview over the total publishing data on printed books. Therefore, the second part of the project involved arranging a fact-sharing seminar where knowledge from the national bibliographic data could be shared and discussed with potential users of the data.

---

[1] Wischenbart, Rüdiger and Holger Ehling, *A methodology to collect international book statistics : framework – indicators – methodology & strategies – groundwork for a test run.* Wien/Frankfurt am Main: UNESCO, 2008, p. 4, Available online: http://unesdoc.unesco.org/images/0018/001824/182475e.pdf, last retrieved 2016-05-30.

[2] *Swedish Institute for Children's Books,* http://sbi.kb.se/en/, last retrieved 2016-05-30 ; See also "Book tasting – some flavours and aromas : children's and YA books published in Sweden 2014", in Statens kulturråd. *New Swedish books: autumn 2015*. Stockholm: Swedish Arts Council; 2015, p. 66-67, Available online: http://www.kulturradet.se/Documents/publikationer/2015/New%20Swidish%20books_final_la%CC%8Ag.pdf, last retrieved 2016-05-30.

No single definition of a book is established as adequate in today's digital environment.[3] The definition formulated by UNESCO at a conference held in Paris, 20 October to 20 November in 1964, concerning an "International Standardization of Statistics Relating to Books and Periodicals" stating that "[a] book is a non-periodical printed publication of at least 49 pages, exclusive of the cover pages, published in the country and made available to the public" is out of date and is not used today in trade statistics.[4] Until 2014 UNESCO:s guidelines for "book production statistics" were used when producing the annual statistical report for the Swedish national bibliography. Generally, there is a lack of standard definitions of basic terms internationally for systematic data collection of international book statistics.[5] In *A Methodology to Collect International Book Statistics* (2008) two reasons are raised as to why there is no hope for any standard definition in sight in the near future. Firstly, there are too extensive culture variations in each country, and secondly, the effects of the digital change that dilutes the clear borders between printed and digital books and, in relation to this, also the fact that the definition of a publisher becomes less clear as self-publishing increases.[6]

In *European Book Cultures: Diversity as a Challenge* (2015) Stephanie Kurschus claims that there is a dual nature of the book that needs to be taken into consideration when looking at the book culture as a whole. The book can be described as both a product and a cultural good.[7] Although it is becoming clearer in many contexts that the book no longer is the most important information medium, the book as a cultural good is still associated with values that traditionally define national bibliographies. The book does not only have the function of being "a technical medium for the storage of information".[8] The discussion of the purpose of a national bibliography can be paralleled with Kurschus' discussion on book culture, where she argues how the book has a strong symbolic value reflecting both individual identity and national identity, even when considering the digital shift in media and the impossibilities in trying to define the increasingly elusive characteristics of the book. Even though it is indisputable today that the national bibliography should include data on all current media types to reflect all intellect output of a country, the traditionally rooted signature of national bibliographies is book data. A book is a book is a book is a book. This allusion to one of Gertrude Stein's most famous quotations, "A rose is a rose is a rose is a rose" can serve as an entrance to the problem discussed in this paper - the future of national bibliographies as sources for data and analysis of national and international book cultures.

There is no ambition in this paper to define what a book is, but in producing national bibliographic data it is important to take into account this dual nature of the publishing output as a whole and re-evaluate the shifting character of the book as both a product and as a cultural good in the digital age. The book is a book is a book is a book, but the soul is not in the medium but in the content, whether it may it be words, sounds or expressed visually in our close interaction with digital devices in our daily lives. In this environment the national

---

[3] Kurschus, Stephanie. *European book cultures: diversity as a challenge*. Wiesbaden: Springer VS, 2015, p. 19,44, ; Kovač M. *Never mind the web: here comes the book*. Oxford: Chandos Publishing; 2008, p. 10-14.
[4] *Recommendation concerning the International standardization of statistics relating to book production and periodical,* Available on line:
http://portal.unesco.org/en/ev.php-URL_ID=13068&URL_DO=DO_TOPIC&URL_SECTION=201.html, last retrieved 2016-05-30. ; Wischenbart and Ehling, 2008, p. 5.
[5] Wischenbart and Ehling, 2008, p. 4.
[6] Wischenbart and Ehling, 2008, p. 5.
[7] Kurschus, 2015, p. 45.
[8] Kurschus, 2015, p. 20.

borders lose effect since it becomes less interesting to limit the processes of reading, writing and publishing with national borders.

2 IDENTIFYING A PUBLISHING PULSE – LIFE STORIES

> *It is true that storytelling reveals meaning without committing the error of defining it, that it brings about consent and reconciliation with things as they really are, and that we may even trust it to contain eventually by implication that last word which we expect from the "day of judgment".[9]*

This quotation is from Hannah Arendt's essay on Isak Dinesen (pseudonym of Karen Blixen), in *Men in Dark Times,* where she also asks in relation to her interpretation of Dinesen's thoughts of storytelling that "no one has a life worth thinking about whose life story cannot be told", which leads to the idea that life could or perhaps ought to be lived as a story and "what one has to do in life is to make the story come true".[10] The national bibliography contains stories that are made accessible through different methods of describing the content accurately as national bibliographic data. To identify a "publishing pulse" in the statistical flow the hypothesis used when developing new methods was that the descriptions are to be analyzed as stories in addition to regarding them as mere metadata. We suggest a shift from metadata to metastories, numerous metastories, i.e. stories that are intertwined with a network of stories crossing individual and national identity. Instead of looking for *one* metanarrative about a country's cultural and intellectual output in one national bibliography, we asked what narrative content was possible to make accessible with visualization experiments. Our starting point was to identify a publishing pulse by analyzing what stories would surface from the national bibliographic data about Swedish publishing trends. We found that it was more interesting to focus on the described data as content and as stories rather than as metadata. In our analyses we found a diversity of stories, memories, thoughts and voices. Diversity is a unique feature of national bibliographic data. Each book is described individually in detail.

What methods could be used for making accessible book data and analysis in conjunction with the precarious future of the publishing market in the digital age? The ambition was to create an interactive tool to make visible what is possible to do with national bibliographic data in order to convey what knowledge can be accessed and to inspire users to explore the data themselves. Broader findings would not be immediately perceivable after only one year, so the goal was to limit the scope to a single theme, and plan for a recurrent, annual event to showcase findings over time. The theme chosen in 2015 focused on "Life writing". Life writing is a broader field than biographies and according to *the Oxford Centre for Life-Writing* it includes: "autobiography, memoirs, letters, diaries, journals (written and documentary), anthropological data, oral testimony, and eye-witness accounts."[11] The theme was chosen because it reflected an obvious trend in society and opened up for including more

---

[9] Arendt, Hannah, *"Isak Dinesen: 1885 – 1963"* in *Men in dark times,* San Diego, Calif: Harcourt, Brace & Company, 1995, p. 105.
[10] Arendt, 1995, p. 105.
[11] Oxford Centre for Life-Writing, "What is Life-Writing?", Available online: https://oxlifewriting.wordpress.com/about/what-is-life-writing/, last retrieved 2016-05-30. ; *King's College London - Centre for Life Writing research*, Available online: http://www.kcl.ac.uk/artshums/ahri/centres/lifewriting/index.aspx, last retrieved 2016-05-30.

media types than the printed books in the future. However, since the annual statistical report today only includes printed books the theme also had this limitation in 2015. The Swedish concept used, "livsberättelser", if directly translated means *life stories*. Therefore the visualization experiments were based on a data set limited by biographical coding where the field of life stories is covered. The choice of theme was also motivated by the core activity of National libraries to ensure the nation's stories for future generations that also are pronounced in the strategic goals of the National Library of Sweden. The National library of Sweden collects, describes, preserves and makes accessible the documented memory of the nation for future generations. But what does memory mean in this context? Memories in national bibliography data could be described as memories shared through stories. What is a story? Not all that is registered in the national bibliography are traditional stories with a narrative, sequential flow. There are dictionaries, cookbooks, atlases, bibliographies that function in a different way. In *So many books* Gabriel Zaid describes these kinds of books; "they don't threaten the recipient with the task of responding to the questions 'Have you read it yet? What did you think of it?'".[12]

   Memory, Life, Language, Stories, Generations, Future, Past, History, Culture, these are only a few examples of words that are used to describe core values of national bibliographies. In Barbara L. Bell's definition of a national bibliography, quoted in IFLA's *Best Practice for National Bibliographic Agencies in a Digital Age*, it is suggested that "a current national bibliography is a mirror that reflects the culture of country".[13] But in this project the mirror metaphor was replaced by a pulse metaphor and instead directed towards *stories* and in particular *life stories.* Culture is regarded as an on-going process rather than as a graspable fixed identity of a country. There is a publishing pulse, the flow can move in many parallel and entangled veins. This means looking at the publishing output not just as a reflection of society but as something taking part in society, a passage of events creating patterns affecting culture rather than reflecting it.

   The stories we found in the statistical flow could be described in biographical terms as born 2015. Perhaps the authors are the parents of these works? All cataloguers who work with national bibliographic data recognize that we register the size of the book. We register how many centimeters and how thick it is by number of pages. In the Swedish national bibliography we do not register weight of the book or gender of the author. Important though is that we register *date and place of birth* of the book. Place of birth is the most crucial criterion that defines if the work is included in the national bibliography or not. Globalization, media convergence and digitalization are factors that make it more complicated to establish data considering place of birth of the media in the future.

3 NATIONAL BIBLIOGRAPHIC LIFE STORIES - VISUALIZATIONS

What unique data within a body of national bibliographic data can be visualized and how? With this project we investigated how we could find methods to visualize what is possible to do with our data by answering questions on *who* publishes, *who* are the stories about, *what* subjects are the stories about, what can be told about time and place, *when* and *where*. These

---

[12] Zaid, Gabriel. *So many books,* London: Sort Of Books, 2004, p 13.

[13] *Best practice for national bibliographies,* http://www.ifla.org/node/7862 , last retrieved 2016-05-30.

are also key features in building a story, *who, what, when* and *where*: a subject, an object, a time and a place. These data elements are available in the national bibliographic data in subject headings, titles and statements of responsibility, classification and information about publishers.

A first step was experimenting with subject headings and titles. We choose a visualization tool called *TagCrowd* to generate a number of "word clouds" that visualized the most common words in titles in a data set limited to life stories.

Fig. 1 *From Abba to Zlatan*.

This initial visualization test already resulted in answers on the questions we posed to the material. Among the most common words in titles for *life stories* we found *I, Life, Mine, Memories, You, We, Where, Sweden, Diary* and *History*. The "word cloud" visualization starts with *ABBA* and ends with *Zlatan*. (Fig. 1).

In the next step a company specializing in visualizations was hired to lead a workshop on our data set limited to life stories. This gave important insight into our continuing work with the project. Why, for whom and what problems did we hope to solve? Our purpose with the visualizations was to communicate and make our data accessible and present the statistics in a manner easy for other to use and interpret. We had the ambition to make statistics on national bibliographic data more accessible by using visualization as a method and to identify what patterns could be found in the national bibliographic data. With the visualizations we furthermore wished to clarify the possibilities of using national bibliographic data as source material for the study and analysis of book culture.

Before starting the visualizations it was important that we clarified what our purpose was and to understand the data in order to being able to draw conclusions. One important key was to offer transparency and documentation of criteria for unique data within a body of national bibliographic data. It has to be clear what a national bibliography is for the users to be able to decide why it is an important data source. But this is also a challenge for national bibliography agencies and needs to be evaluated and developed over time. It was furhtermore

important to limit the scope of the material and make priorities. This led us to a method for visualization performed in four steps: 1. Define a purpose with the visualization, 2. Explore the content of the data, 3 Produce the visualization, 4. Evaluate.

The production of the visualizations was performed by testing a number of free visualization tool software products. The software chosen was Tableau Public.[14] Data sets were downloaded from the library system into MS Excel for initial exploration prior to beginning work with the visualizations. At this stage the data sets were examined and adjusted to fit our purpose before loading them to Tableau Public. In Tableau Public "a story" of visualized life stories from the Swedish national bibliographic data of 2015 was constructed. In these visualizations the users are able to explore subject headings, follow a map on geographical subject headings, compare date of birth of the authors, explore number of titles by each publisher and learn more about how the classification codes for life stories are built. In these visualizations of life stories knowledge is revealed about the national bibliographic data in colours, sizes, location in a manner that would not be possible to access in the regular library catalogue. Another advantage was that the user could interact with the data and create their own data sets. (Fig. 2)
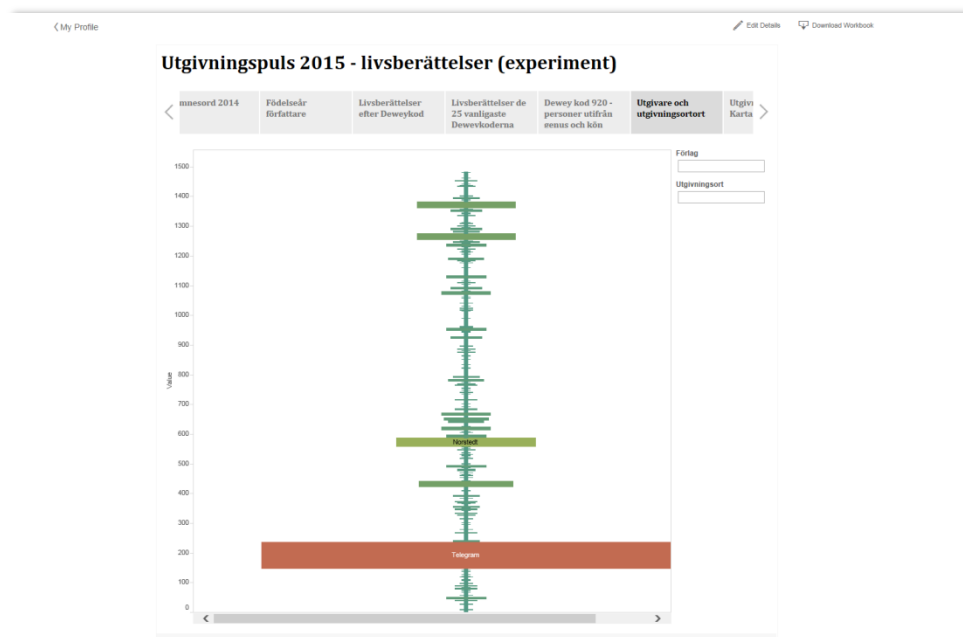

Fig. 2 Visualization experiment Life Stories (Livsberättelser), *Number of publishers.*

---

[14] https://public.tableau.com/, last retrieved 2016-05-30.

4 CONCLUSIONS

The visualizations that resulted from the experiments on the limited data set of life stories 2015 were a first step for the National Library of Sweden when continuing to develop methods to increase the transparency on type of data registered. And as an effect of this visualization process thereby enhance the accessibility of national bibliographic data for the users. The project also resulted in an improved knowledge for some problematic features of the data. The continuing plan for the project will therefore involve users when developing the methods for making this data accessible. One issue related to the uniqueness of national bibliographic data is the level of granularity when it comes to categorisations of the books, formulated in subject headings and library specific classification systems.[15] However, these systems are developed based on the concept and physicality of the printed book, where it is important that the bibliographic record will point to a particular item placed on a specific location on a physical shelf in the library. One result from the visualizations related to this issue was that it was clear that the data was characterized by diversity. For example, most classification codes and subject headings were used only once to describe one single title. Diversity is a unique feature of national bibliographic data. In our analyses we found a diversity of stories, memories, thoughts and voices. Each book is described individually in detail. This property of the data has to be considered when rethinking models of cataloguing, in the case of the National Library of Sweden, the on-going transition into the new cataloguing rules (RDA). Is the granularity regarding the descriptions of subjects necessary for the users to get access to the information today and in the future? One answer to this is that exactly the accuracy of the data is one of the strengths of national bibliographic data in comparison to other data sources on the publishing output. A second answer to this question, drawn from the results of the visualizations, is that for the continuation of the project we have come to the conclusion that we will have to investigate clustering methods to analyse subjects on a more overall level to be able to group data that is registered in various formats. In future years, we will therefore focus on developing methods for clustering adjacent data sets and analyzing how they are related. Taking this path also opens up for international perspectives on how to make national bibliographic data accessible as a way of developing international book culture statistics and follows technical models of making the data accessible through linked open data.

---

[15] Wischenbart and Ehling, 2008, p. 11.

References

Appelman, Marja, and Marcel Canoy. "Notes and Communications, Horses for Courses: Why Europe Should Not Harmonise Its Book Policies." *De Economist : Quarterly Review of the Royal Netherlands Economic Association*. 150.5 (2002): 583-600.

Arendt, Hannah. *Men in Dark Times*. San Diego, Calif: Harcourt, Brace & Company, 1995.

Elkins, James. *Visual Literacy*. New York: Routledge, 2008.

Eurostat, *Cultural Statistics*. Luxembourg: Office for Official Publications of the European Communities, 2011.

Kovač M. *Never Mind the Web: Here Comes the Book*. Oxford: Chandos Publishing; 2008.

Kurschus, Stephanie. *European Book Cultures: Diversity as a Challenge*. Wiesbaden: Springer VS, 2015.

Manovich, Lev. "What Is Visualisation?" *Visual Studies*. 26.1 (2011): 36-49.

Mitchell, W J. T. *On Narrative*. Chicago : Univ. of Chicago Press, 1990.

Recommendation concerning the International Standardization of Statistics Relating to Book Production and Periodical, Available on line:
http://portal.unesco.org/en/ev.php-URL_ID=13068&URL_DO=DO_TOPIC&URL_SECTION=201.html

Statens kulturråd. *New Swedish Books: Autumn 2015*. Stockholm: Swedish Arts Council; 2015, p. 66-67, Available online:
http://www.kulturradet.se/Documents/publikationer/2015/New%20Swidish%20books_final_la%CC%8Ag.pdf, last retrieved 2016-05-30.

Statistical yearbook, UNESCO, Paris, 1964-1999.

UNESCO. Office of statistics, *An International Survey of Book Production During the Last Decades*, UNESCO, Paris, 1982.

Wischenbart, Rüdiger and Holger Ehling, *A Methodology to Collect International Book Statistics : Framework – Indicators – Methodology & Strategies – Groundwork for a Test Run.* Wien/Frankfurt am Main: UNESCO, 2008, Available online:
http://unesdoc.unesco.org/images/0018/001824/182475e.pdf, last retrieved 2016-05-30.

Zaid, Gabriel. *So Many Books*. London: Sort Of Books, 2004, p 13.

Zell, H. M. "How Many Books are Published in Africa? The Need for More Reliable Statistics." *The African Book Publishing Record*. 39.4 (2013): 397-406.