Opening up the bibliographies for the future – A collaborative research-driven model for bibliographies

Hege Stensrud Høsøien

Scholarship and Collections, National Library of Norway, Oslo, Norway
Hege.Hosoien@nb.no

Irene Hole

Acquisition and Cataloguing, National Library of Norway, Oslo, Norway
Irene.Hole@nb.no

Abstract:

Although the production and mediation of various kinds of bibliographical data represent one of the library's core activities, today a number of practices, technologies and disciplines influence this vast field. Even though bibliography is usually understood as a product of library science models, the field also includes certain kinds of empirically-based research, philology and book history, among other disciplines.

Within the library community there has been a focus on the present situation of constant technological changes, the emergence of new search technologies/user-interfaces, merging catalogues with bibliographies, integrating individual library catalogues into meta-catalogues, and large-scale digitization programs. Another focus has been the realization and implementation of FRBR, linked data and semantic technologies.

Until recently the focus on special bibliography production has been to strengthen the national research infrastructure. Today, however, special bibliographies also seek to incorporate research data and theoretical models from the academic community.

This paper discusses the National Library of Norway's model for producing special bibliographies. Examples from three special bibliographies, Hamsun, Bjørnson, and Solstad, illustrate current aspirations for special bibliographies: to be research-driven, collaborative and technologically dynamic.

Opening up the bibliographies for the future – A collaborative research-driven model for bibliographies

Addressing the challenge put forward by the IFLA committee and by the title of this session, **Opening up the bibliographies for the future**, it seems only fair to ask, before we start thinking about how to open bibliographies for the future: What is bibliography today, to us, to librarians and libraries? The word bibliography itself comes from the Greek βιβλιογραφία, *bibliographia*, literally "book writing". Greek writers first used the word to describe the copying of books by hand. In the 12th century, bibliography started being used for "the intellectual activity of composing books", but it was not until the 17th century that "the description of books" emerged as the modern meaning of the word.

Reflecting upon the title for this session, **Opening up the bibliography for the future** I realize that the title itself and the use of "the bibliography", in its singular, particular form, is, in fact, just what I would like to address. I think an earlier title for this session was **Bibliographies for the future**. Both titles reflect an understanding of the word bibliography as a noun, a product, or products. Perhaps also an understanding of the word that reflects the bibliographic form or convention we are most accustomed to meeting in the library, the enumerative bibliography, the organized list of books, for example a National Bibliography. Speaking from a position within the library it is easy to see that we surround ourselves with, immerse ourselves in bibliographies and bibliographical activities: registering and mediating bibliographical metadata, discussing their standards and rules, speaking MARC and MODS, Dublin Core and RDF.

Meanwhile, at the universities and in the academic communities, more attention has been given to another form of bibliography, the descriptive bibliography (and its various manifestations, analytical, historical etc.). This form is the systematic and often very minute description of books as physical objects, of interest to philologists of various kinds, and increasingly interesting to the academic community engaged in the interdisciplinary field of what we may call histoire du livre or book history.

In a foreword to his influential essay *Bibliography and the sociology of texts* (McKenzie: 1984), the New Zealand-born bibliographer, D. F. McKenzie (1931–1999), later reader and professor of bibliography and textual criticism at the University of Oxford, states:

> "The familiar historical processes by which, over the centuries, texts have changed their form and content have now accelerated to a degree which makes the definition and location of textual authority barely possible in the old style. Professional librarians, under pressure from irresistible technological and social changes, are redefining their discipline in order to describe, house and access sounds, static and moving images with or without words, and a flow of computer-stored information. By contrast, academic bibliography has only recently begun to find fresh stimulus in those developments and to tap the new experience and interests of students for whom books represent only one form of text." (McKenzie: 1984, 1)

In a few sentences McKenzie says several things worth pondering. Speaking of bibliography, McKenzie includes both the enumerative and the descriptive type, but views bibliography not so much as a product than as a discipline, or disciplines. He describes how the process by which texts change their form and content has accelerated – a point anyone working within a library can easily relate to – and how librarians are redefining their discipline in order to "describe, house and access sounds, static and moving images with or without words, and a flow of computer-stored information". This reflects a widening of the scope of what we traditionally are used to viewing as the object for bibliography, namely the book, or at least written documents of some sort.

As a contrast to the professional librarian – eager to redefine her discipline in the face of "irresistible" technological changes – McKenzie poses academic bibliography that has only recently "begun to find fresh stimulus in those developments". Identifying two different bibliographical disciplines, one of library science, and one of academic bibliography, it seems only logical that the two schools would have different perceptions, of not just what constitutes bibliography as a product, or as a discipline, but also how bibliography can be "opened up for the future", as is the theme for our session.

To McKenzie the consequence of the accelerated pace of changes to textual form and content is that the definition and location of textual authority is "barely possible in the old style". His comment must be understood in relation to the kind of bibliographical discipline where edition philologists seek to establish textual authority by trying to understand the author's intention and to establish which text can serve as copy text. The motive is to determine which text is to be trusted as the original – to which other texts may be compared and new editions published – be this text the manuscript the author delivered to the printing house, the first published edition or the last edition published where the author was involved in the editorial process.

Opposing solely focusing on establishing textual authority as the primary goal of descriptive bibliography, McKenzie introduces his notion of bibliography as a sociology of texts:

> "[…] bibliography is the discipline that studies texts as recorded forms, and the processes of their transmission, including their production and reception […] I define 'text' to include verbal, visual, oral, and numeric data […]" (McKenzie, 1984: 12).

This shift in focus within academic bibliography, from philological concerns with textual authority to the "sociology of texts", was propelled by bibliographers and book historians like McKenzie himself, but also by the likes of Philip Gaskell, Robert Darnton and Roger Chartier, to mention a few of the best known. One could say that the new focus constituted a paradigmatic shift in academic bibliography and one that has served to revitalize the field considerably.

Within the library community, on the other hand, the expectations for the future of bibliography, in addition to extending the notion of texts to include verbal, visual, oral, and numeric data, seem to focus primarily on practical and technological advances: The emergence of new search technologies, user-interfaces and

applications; the integration of catalogues with bibliographies; the merging of individual library catalogues into meta-catalogues and the implementation of new models for representing metadata like RDF (Resource Description Framework). And as anyone attending a large library conference the past few years can report: we dream about the realization and implementation of FRBR, (Functional Requirements For bibliographic Records), linked data and semantic technologies.

McKenzie wrote his Bibliography and the sociology of texts in 1984, almost 30 years ago. When we started working with our new model for producing certain kinds of bibliographies in the National Library of Norway a few years ago, the situation was pretty much the same. While librarians were cataloguing and producing bibliographical metadata and bibliographies like the National Bibliography, professionals in the academic communities were producing bibliographies in Word, and publishing them as books. Moving closer to the turn of the century we started converting the Word files into HTML and publishing them on websites. Particularly if there was a large corpus of texts being handled, we started punching the data into a proprietary database, using a lot of labor from librarians.

Deciding to do things differently in the future, we started thinking about a new model for producing certain bibliographies in the National Library, a model that would take the best from both bibliographical traditions. The model was a collaboration between the National Library's departments of research, cataloguing and IT. We wanted a digital solution that was pragmatic, technologically dynamic and sustainable and that could be used both by the National Library and other libraries. After considering different options we decided to produce bibliographies inside BIBSYS, Norway's collaborative meta-catalogue that unites more than 120 large libraries, among them all the University libraries, most of the research libraries and the National Library. Using BIBSYS as a tool for the production of bibliographies, we would then give each new bibliography an individual user interface on our own webpages. Furthermore, using BIBSYS gave us significant advantages in comparison with other possible solutions. It gave us the opportunity to use MARC (Machine Readable Cataloguing), a well-known international standard for bibliographical records that makes it relatively easy to cooperate with others, map and migrate data and so on. BIBSYS holds data about a large number of library collections and their individual titles and copies, and this gave us the opportunity to identify a great number of relevant documents for inclusion in the bibliography. After identifying them, we could recirculate high-quality metadata from the National bibliography (BIBSYS holds the National bibliography since 2011) and other libraries by tagging already-existing posts to the bibliography. BIBSYS also holds a virtual authority register for names, developed and managed in cooperation with the National Library, and so choosing BIBSYS gave us the opportunity to draw on and contribute to the national authority register for names. The integrated solution gave users information on where to find a copy of a book, and access to loan options.

Finally, and this would prove to be increasingly important: BIBSYS holds the data and the links to the digitized files from the National Library's digitization program. Started a few years earlier, the digitization program, aiming at digitizing the National Library's entire collection, was gaining momentum, and the government was eager to increase the rate of digitization and access. In 2009 a radical extended collective licensing agreement called Bokhylla (The Bookshelf) between the National Library and the organization Kopinor, representing authors, publishers and other rightholders,

gave all users with a Norwegian ip-address free access to 50.000 Norwegian book titles published in the 1990ties. In 2012 the contract was positively evaluated by both parties and the agreement was extended to printed books published in Norway until and including the year 2000. In 2013 110.00 books are digitized and freely available, and the agreement allows for 250.000 titles by 2017. While working through a corpus for a bibliography all titles have been digitized.

Within the new model, the Hamsun bibliography was the first that the National Library made. As part of the 150-year national celebration, the Library presented a bibliography on Norwegian Nobel Prize winner Knut Hamsun (1859–1952). The bibliography was produced in cooperation with the University Library of Tromsø and proved the model's potential for collaborative bibliographical production. Upgrading and complementing already-existing bibliographical posts, our work enhanced the collaborative meta-catalogue and the National Bibliography, as well as added many new entries to the national authority register of names for persons and institutions. Working systematically through a textual corpus we realized the model's potential for identifying lacunas in the National Library's collection, documents that had gone missing, unnoticed, over the years, and documents that had in fact never been legally deposited for one reason or another.

Gaining confidence and competence we went on to produce a bibliography on Norwegian Nobel Prize winner Bjørnstjerne Bjørnson (1832–1910) as part of the national commemoration of 100 years since his death. Freeing ourselves from previous library routines, and thinking now of bibliography not only as infrastructure for research but as a research collaboration and ultimately research mediation and dissemination, we worked closely with researchers and Bjørnson's biographer Edvard Hoem. Changing routines of working through a corpus alphabetically or chronological, we adapted pragmatism as our new modus and prioritized working wherever our work was most needed. We added material genres researchers were interested in, such as newspaper articles and published speeches, identified documents they were looking for and made them available, adding layers of new metadata. For example, we added tags for geographical location on a map, and adapted and exported data to a Google maps application, making it possible to visualize Bjørnson's movements and engagement. In return, researchers contributed their specialized knowledge by identifying authors and pseudonyms, suggesting key words, identifying relevant documents etc. Thus, researchers contributed not only user-generated metadata; they also yielded research-generated metadata. Their insights enriched the National Bibliography, the National Authority Register and the collaborative catalogue.

Working on Hamsun and Bjørnson the textual corpuses were large and we made strict appraisals of what to include. In 2011, however, we started working on a bibliography to celebrate one of Norway's most popular contemporary authors, Dag Solstad (1941–). For the first time we could attempt to including everything written by and about an author. Taking a closer look at an authorship made it clear that this model for producing bibliographies had one serious limitation: it did not hold any information on the relations between the documents, which made it difficult to gain an overview and orientation the way researchers and the general public prefer. Bibliographical standards from the time when bibliographies were published as books, and when records of library holdings existed in card catalogues, simple functionality like

following how one work was published in many different editions over the years are abandoned in the digital library catalogue. Adding new layers of metadata and working on the search algorithm and the user interface, we were able to recreate some of these relations and give the user a way of navigating through an authorship (the works in the many editions and translations, and finally the reception). This was a partial and pragmatic realization of some of the principles that underlie FRBR.

Working our way through a textual corpus of works and its manifestations also made us realize that although we had utilized the bibliographical models usefulness for identifying lacunas in our collections, we were now able to identify lacunas of another sort. Over the past few years, the model has allowed us to build up our competence on book history and the materiality of books. Consequently, we were able to sort through the corpus, give attention to different bindings of the same edition, note cases where the library held a book, not in the original binding but a library binding, and identify books with interesting provenance. Thus, we were able to move closer to fulfilling our ambition of holding a complete collection of Norwegian prints in all their material, physical and philological variations.

We continue to use our bibliographical model as a tool for development. Our plans for the future are to work with creating an authority register for works. Together with the authority register for names, already published in VIAF and as linked data, this development will take us one step closer to utilizing sematic technologies. Moreover, we are currently working with a filmography, along the same lines. Metadata pertaining to films are held in our other large metadata repository, MAVIS, which also holds metadata for photography, radio, television and music. Joining these two repositories and giving datasets a user interface will enable us to make truly multimedia bibliographies in the future.

Finally, we are also aware that bibliography may play a significant (if small) role in the future of digital humanities and culturomics. Schreibman et al. in *A companion to digital humanities*, describes the situation:

> "The World Wide Web spurred a generation of pseudo-publication: documents more broadly available than any print publication in history could at any given time reach millions of machines. The same documents often ran, however, under individual accounts, with many URLs being changed or pointing to documents that were no longer online or, arguably worse, that had been substantively changed since the original link had been added. A variety of library repositories are now coming into use." (Schreibman et al.: 2004)

As libraries digitize their collections and researchers shift their focus from the close reading of a few canonized texts to the distant reading of enormous digital libraries, as Franco Moretti describes, the demand for digital textual corpuses that come with reliable metadata will increase. The libraries can fill the role of trusted repositories, guaranteeing not just the quality of the digitization, but that the digitization can be trusted to represent precisely what the user is seeking, for example the first edition of Knut Hamsun's Sult, not an arbitrary later edition or abbreviation (Moretti: 2011). Considering this new role for the libraries, we might be witnessing yet another shift in bibliography: from the original question of textual authority, to the question of textual sociology and back again to a new form of textual authority.

Let us return now to the challenge in the session title: Opening up the bibliographies for the future. We can answer that we may extend our notion of text, to look at both enumerative and descriptive bibliographies – at library science and academic bibliography – as one large interdisciplinary field. We can answer that it is beneficial to invite research data and theoretical models from the academic community into the library machine.

Bibliography

Bjørnstjerne Bjørnson bibliography. Newspaper articles and speeches by Bjørnson; literature on Bjørnson http://www.nb.no/bibliografi/bjornson/search.jsf

Hamsun bibliography : literature on Knut Hamsun
http://www.nb.no/bibliografi/hamsun/search.jsf

McKenzie, D.F. 1999. *Bibliography and the sociology of texts*. Cambridge University Press.

Moretti, Franco. 2011. *Network theory, plot analysis*. Palo Alto: Stanford University. (Pamphlet/Literary Lab; 2)

Schreibman, Susan; Siemens, Ray; Unsworth, John. 2004. *A companion to digital humanities*. Blackwell.

Solstad bibliography : literature by and on Dag Solstad
http://www.nb.no/bibliografi/solstad/search.jsf