The online Bach Bibliography:
a resource maintained by the global network of scholars

A paper read at IAML conference in Tallinn on 8 July 2003

Yo Tomita

Queen's University Belfast

INTRODUCTION

For a majority of musicologists, bibliography is not their main subject of research, but an essential tool, facilitating quick access to information. It is important that we, musicologists, can find relevant literature as soon as possible, and that we do not overlook any significant research by other scholars in the field. Considerable effort is needed to acquire a comprehensive bibliographic knowledge even in a specialised area of research. This can be a laborious process, but modern on-line resources have the power to speed up our searches and ensure their comprehensiveness.

The Bach Bibliography (http://www.music.qub.ac.uk/tomita/bachbib/) is an online resource for those who study the life and works of Johann Sebastian Bach. Primarily it facilitates the advanced study of the composer who has been discussed and seriously scrutinized from many different angles for more than two centuries. To meet the complex needs of scholars, it offers a range of unique features, e.g. closely-integrated crossreferences, highly-structured search options, the inclusion of unpublished papers, thesis, and the online review of the hot-from-the-press books. It went online in May 1997, which quickly attracted the interest of Bach scholars all over the world. Presently, it is accessed roughly 200 times per day. While the maintenance of both the database and the website is my task, I also receive assistance from about seventy scholars who periodically provide new information about their own publications as well as those of others that escaped my notice.

My paper today concentrates on the following four issues:

1. BACKGROUND of its data resource, where I discuss the origin of my database, how it began, and how it reached to the present form;
2. SEARCH STRATEGIES, where I discuss how they are developed and implemented so that the database functions for scholars in a useful way;
3. UNDER THE BONNET, in which I discuss some technical details such as the search engines and data format;
4. FUTURE PLANS, where I am going to outline my vision of developing a completely new, multi-user management environment, the ways to expand data set (to include 19C journals), seeking a more streamlined collection of new data, and implementing improved search techniques using the data-mining concept.

BACKGROUND

Johann Sebastian Bach (1685–1750) is perhaps the most intensively researched composer of Western music. This is clearly reflected in the flood of new publications that deal with his life and works. Taking the tercentenary year of 1985 as an example, there are 782 references that are currently known to me. For the quarter millennium, celebrating his death in year 2000, the number of records has already reached 632, and this figure is expected to grow as the information is still being gathered (see Figure 1). Bach research is becoming increasingly specialised and fragmented to the extent that even an established scholar can find it hard to keep up with the latest developments.
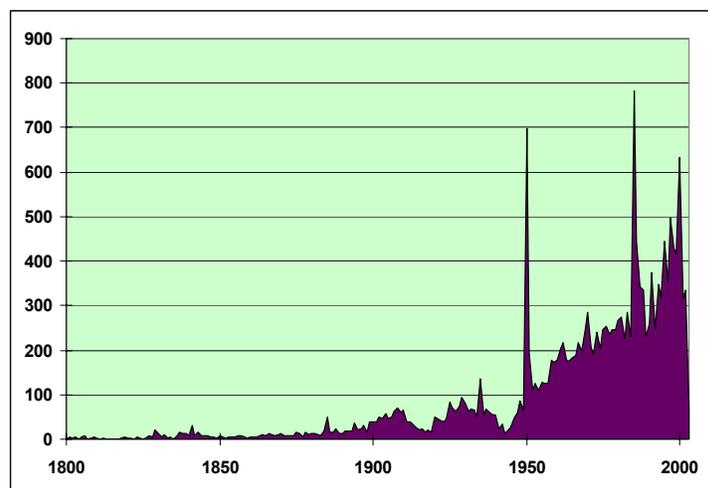
Figure 1: Statistical breakdown of Bach references from 1800 and 2003

Bibliographical studies on J. S. Bach and his works have a long and proud history. With the establishment of the *Neue-Bachgesellschaft* in 1900 and its journal the *Bach-Jahrbuch* from 1904, the information was collected, classified and published periodically, as summarised in Table 1.

Table 1. The coverage of 'Bach-Schrifttums' published in *Bach-Jahrbuch*

Coverage	Items	*Bach-Jahrbuch*	Compiler	Notes
1800–1904	453	2 (1905) †	Max Schneider	mostly German literatures; no page numbers given
1905–1910	321	7 (1910) †	Max Schneider	some with page numbers
1910–1944	299	(see notes)	Christoph Wolff	in *Bach-Bibliographie* (1985)
1945–1952	941	40 (1953) †	Wolfgang Schmieder	more systematically described
1953–1957	402	45 (1958) †	Wolfgang Schmieder	
1958–1962	807	53 (1967) †	Erhard Franke	
1963–1967	715	59 (1973) †	Rosemarie Nestle	
1968–1972	884	62 (1976) †	Rosemarie Nestle	
1973–1977	845	66 (1980) †	Rosemarie Nestle	
1978–1980	560	70 (1984) †	Rosemarie Nestle	
1981–1985	1144	75 (1989)	Rosemarie Nestle	
1986–1990	1174	80 (1994)	Rosemarie Nestle	
1991–1995	1331	86 (2000)	Karin Germerdonk	
[1996–2000	c.1500	91 (2005)]	?	projection based on my data shown in Figure 1

† also repr. in *Bach-Bibliographie*, ed. Christoph Wolff (Kassel, 1985)

The first of these was published in the second issue of the *Bach-Jahrbuch* (1905).[1] In it was listed 453 Bach references,[2] mainly of German sources that appeared between 1800 and 1904. Five years later, another list was published; it consisted of those older references missing in the earlier listing, recently published works and reprints. From these emerged the tradition of publishing a new bibliography every five years. No lists were published from the period 1910–1944 but a list was subsequently compiled by Christoph Wolff and published in the tercentenary year, in which he reprinted all the previous articles issued since 1904 with a newly compiled index of authors.[3]

---

[1] Max Schneider, 'Verzeichnis der bischer erschienenen Literatur über Johann Sebastian Bach', *Bach-Jahrbuch*, ii (1905), 76–110.
[2] The count excludes 28 entries of Bach's portraits and busts, listed under 'Anhang', 109–110.
[3] *Bach-Bibliographie. Nachdruck der Verzeichnisse des Schrifttums über Johann Sebastian Bach (Bach-Jahrbuch 1905–1984). Mit einem Supplement und Register*, ed. Christoph Wolff (Kassel, 1985).

In 1950—the bicentenary year of Bach's death—there was an important development, not only for Bach scholarship but also for bibliography in general. In that year appeared Wolfgang Schmieder's *Bach-Werke-Verzeichnis*, which revolutionised the way we refer to Bach's works.[4] His involvement with the *Bibliographie des Musikschrifttums* from 1950 onwards also resulted in significant changes in the way the information of bibliographical material is dealt with; he advanced not only the standard of precision in describing the literature but also the system in which they are classified.[5] This is clearly reflected in his 'Bach Schrifttums' articles published in the fortieth and forty-fifth issues of the *Bach-Jahrbuch*.

The most recent list was published in the eighty-sixth issue of the *Bach Jahrbuch* (2000), which covers the literatures published between 1991 and 1995. It can be predicted that the number of entries is set to rise to a record level when the ninety-first issue of *Bach-Jahrbuch* (2005) publishes the list of literatures including those published during the millennium year. It is doubtful whether we will ever see again a single-volume reprint in a book form of the kind published by Wolff in 1985. Needless to say, it is impractical to manage the sheer quantity of information in this way. Electronic publication is therefore much better suited to this type of project, for we can maintain quickly and cost-effectively the most up-to-date information, while at the same time benefiting from advanced search facilities to sift through the vast amount of data.

THE ON-LINE BACH BIBLIOGRAPHY AND ITS STRENGTH

The data that form the basis of my Bach Bibliography are collected from the above-mentioned series of printed resource. I thus cannot claim that this online resource is my own property, for it is the heritage of the community of Bach scholars of past and present. I am therefore pleased and assured to receive ongoing contributions from so many individuals as well as support of various kinds from Bach organizations, in particular the American Bach Society, the Riemenschneider Bach Institute and the Bach-Archiv Leipzig.

To expand the data collection, I also invested my time heavily. In search of more information, I examined numerous books and articles as well as other well-established bibliographic resources, such as the *Bibliographie des Musikschrifttums* and RILM; I also spent many hours in the dark corners of research libraries, scanning through periodicals looking for hitherto-unnoticed articles. The checking of multiple resources was useful in identifying typographical errors and other irregularities in the data. Through this process I also learned a great deal about the subject such as the changing trends and focuses of research and debate over the years, and who have been working on what subjects, etc.

Today my Bach Bibliography contains 19,029 records of bibliographical references that are considered useful for the scholarly discussions of Bach's life and works, which is, as far as I am aware, the largest bibliographic resource for this purpose. This one can compare with some of the best known resources, as shown in Table 2.

Table 2. Comparing the scope of bibliographical data on 'J. S. Bach'

Bibliographical Reference	items	coverage	notes
The on-line Bach Bibliography	18300	1800–2003	plus *c*.700 of pre-1800 refs
'Bach-Schrifttums' series in *Bach-Jahrbuch*	9876	1800–1995	see Table 1 for details
rilm online (via OCLC FirstSearch)	7857	1967–2003	keyword search 'Bach'
IIMP CD-ROM (reviewed in 1998)	1588	1920–1995	keyword search 'Bach'
ArticleFirst (via OCLC FirstSearch)	629	1990–2003	keyword search 'Bach'

---

[4] Wolfgang Schmieder, *Thematisch-systematisches Verzeichnis der musikalischen Werke von Johann Sebastian Bach. Bach-Werke-Verzeichnis <BWV>* (Wiesbaden und Leipzig, 1950).
[5] *Bibliographie des Musikschrifttums*, hrsg. im Auftrage des Instituts für Musikforschung Berlin von W. Schmieder. 1950/51f. (Frankfurt am Main, 1954 f).

SEARCH STRATEGIRES

The Bach Bibliography offers many unique facilities that greatly enhance the speed of information retrieval:

- With 'Essential Collections', you can check both the list of contents and bibliographical details of reference editions and important series and periodicals, as if you are picking up a book in your own virtual library;
- With 'New Publications', you will learn the news of latest and forthcoming publications, as if you are in a bookshop.
- With the 'On-line Review' of newly published books, which I started in January 2000, you can find out if you will need to get hold of a copy for your own work.

It also offers conventional search facilities with keywords. While the 'Simple Search' will suffice if you can think of some keywords, the 'Complex Search' will perform wonders as long as you know what you are looking for. With it you can short-list the literature by specifying the genre, specific musical works, the aspects of discussion, language used, the types of publication, or the combination of these with the click of the mouse. So, you can now search for literature that deals exclusively with the 'analytical' discussion of the 'Well-Tempered Clavier' or a 'theological' inquest into the 'St Matthew Passion' in English.

When the result of your search is displayed, you can do several things:

1. view further details by clicking on the title in order to see additional information about the publication;
2. list the author's other publications by clicking on his or her name, which might highlight related publications by the same author;
3. search all the other articles in the same volume by clicking the 'series', which may show the broader context in which the article was published.

UNDER THE BONNET

The database was originally compiled using the BBC Micro in mid 1980s, then migrated to dBase IV on an IBM compatible PC before it settled on the program called "Idealist", a text-based database software running on Microsoft Windows. The system actually consists of two databases, the main database of references and its child that handles the abbreviation of periodicals. Currently I have no plan to use any other database software.

The selection of data for my Bach Bibliography was determined by my own self interest. The guidelines given in Table 3 have been used for the selection of references:

Table 3. Selection criteria of references for the Bach Bibliography

Selected references	Excluded references
• Published books of scholarly contents and reviews[6]	• Printed editions of music and scores (except those of a highly scholarly nature)
• Published articles of scholarly contents (including reviews and correspondence)	• Recording and its reviews (except accompanying books or booklets of a highly scholarly nature)
• Published facsimiles	
• Unpublished dissertations (Master and Doctor) from universities	• Books for children
• Unpublished papers read at recognised conference or meeting[7]	• Fiction
	• On-line essays posted on the internet

This could mean that for those scholars who are specialised in the field of music publishing, discography, modern performance practice, primary school education and the entertainment industry, my Bach Bibliography may not be very useful. Clearly, I would have to reconsider whether or not I should include the references for the above-mentioned aspects of research if the aim is to embrace all areas of Bach scholarship.

Some of the unique features of the Bach Bibliography, such as the above-mentioned 'Complex search', are achieved by the inclusion of some additional fields in the database as shown in Table 4.

Table 4. Additional fields made available in each record

Field	Notes
Genre	general classification of genre, i.e. **v**ocal, **k**eyboard, **i**nstrumental, etc.
Music	specifying Bach's works (in 3-letter abbreviation) discussed
Aspect	specifying the aspect of discussion in 3-letter abbreviation
Publication type	**b**ook, **a**rticle, **t**hesis, **f**acsimile, **m**usic, unpublished **p**aper, **r**eview
Language used	specifying with 3-digit telephone codes: 001=English; 049=German; 033=Dutch...
Reviews	list of its reviews or author's own summary in separate publication
Created	date on which the record was first created
Modified	date on which the record was last modified
Memo	reserved for future use for feedback, comments, etc.

Of these, the fields called 'Music' and 'Aspect' merit special attention, for they make the Bach Bibliography both unique and useful. This definition and classification are the result of my humble attempt to describe the kind of ranges and diversities of discussion that Bach Studies enjoy, and for this reason I did not adapt the generic system that was developed by librarians. Still, I have to admit that while this system works fine with majority of references, it does not work well with the others where the classifying of the references in this way is somewhat arbitrary or difficult, owing to the varying degree of the focus or scope of discussion. In fact, significant portion of my information come from secondary source, and it is not infrequent that I find it hard to fill in these index fields when the title does not sufficiently reflect the contents of discussion, resulting in these fields left blank for the moment.[8]

The search engines are written in *perl* script by Malcolm Vincent at my request, who then worked for the Computer Services at the Queen's University Belfast. The programs are both portable and platform independent, which process the data mounted on the webserver that are formatted in comma-delimited text file. We are happy to make them available for other scholars to launch a similar database project: so

---

[6] Reviews are normally included as a part of the "referred" reference item under "review" field. Exceptions may apply if the review is either unusually substantial or significant.

[7] When it is published, the entry in the database is also modified as published article.

[8] This tendency is particularly strong in less serious, magazines-type series, such as the *Musical Times*. Abstracts given in *rilm* are very useful in this regard.

far to date, the "International Inventory of Villancico Texts", the database compiled by Dr. Paul Laird of University of Kansas (http://www.sun.rhbnc.ac.uk/Music/ILM/IIVT/), is the only project that made use of these scripts.

FUTURE PLANS

As for its future development, I have a number of interesting ideas.

The most important is the question of DATA COLLECTION and EDITORIAL CONTROL, as this project cannot be sustained in the current form forever where I do most of these tasks on my own.

As for dealing with new or relatively recently published materials, the best solution seems to be to invite scholars to add or modify records directly online, who then assume the responsibility for both checking the accuracy and maintaining the integrity of these records, including the index entries. Likewise the monitoring of journals can be entrusted to them.

The collection of older materials is equally important. They are yet to be searched thoroughly. This is primarily due to the fact that such sources are rare and only available in certain research libraries, some with relatively limited accessibility. The recent launch of RIPM is thus a very welcome addition to our research resources. Given the rapid growth of interest in recent years in researching reception history and "the Bach revival movement" in the nineteenth century, serious bibliographic coverage of this period is long overdue.

Although the on-going RIPM project will surely recover many important references from this period that were forgotten, what we need also is a critical bibliography of all the articles published in newspaper and magazines, not only those substantial ones but also the advertisements and reviews of the early editions and the concerts of Bach's music. With the arrival of digital age where many resources are being digitised and made available for online analysis, the bibliographical databases will predictably play an integral part of a comprehensive e-Science system.

Another area requiring improvement is the search engine, which affects the performance of a database system. There is considerable potential for improving the search techniques currently used by the Bach Bibliography.[9] While it already provides some of the useful facilities outlined above, it does not offer a number of other functions, such as the ability to sort the search result alphabetically, chronologically or in the order of relevance to the query, or to narrow or widen the search. As the data grow larger, it will become more important to implement not only these functions mentioned above but also more effective searches driven by Artificial Intelligence techniques. More intuitive and powerful facilities that emulate the way experts would react to the bibliographical information, namely how they predict which literatures are more important than the others, are now possible. In order for an AI program to provide effective answers, we need to consider a range of issues that form the basis of its knowledge. These are tentatively listed in Table 5.

---

[9] The present search engine was developed in 1997 by Malcolm Vincent of the Computer-Based Learning Unit at Queen's University Belfast, to whom I am very grateful.

Table 5. Knowledge required for measuring the significance of a scholarly contribution by an AI-driven search

category of knowledge	about the author	about the book	about the article
knowledge extractable from the existing database	• publication record • participation in NBA project	• publisher's prestige by its history of continuity • page numbers • year of publication • number of reviews	• prestige of the journal by its history of continuity • page numbers
knowledge to be assigned (either manually or by vote on the internet) into database	• specialisation • education • posts held	• specific scholarly contributions • the number of citation by others	• prestige of the journal by referee status • specific scholarly contributions • the number of citation by others

FUTURE DIRECTIONS AND THE QUESTION OF INTERNATIONAL COLLABORATION

While some of the ideas I outlined above are already been investigated by my students, there is a limit to which one can do and achieve. In order to maintain its usefulness while coping with the ever growing size of the data, a good will of scholars alone may not be sufficient to sustain the project. I may have to find proper resource sooner or later. There is an alternative solution if there is a political will: we should give a serious thought on establishing a dynamic distributed research resource network, in which RILM, RIPM and other commercially managed resources, play an integral part. In doing so, many other small-scale, specialized projects such as mine can be created, without duplication of efforts. I welcome your ideas and suggestions.