COLLABORATING WITH THE COLD REGIONS BIBLIOGRAPHY PROJECT: THE *SPRI* EXPERIENCE

Hilary Shibata
Scott Polar Research Institute
Cambridge

Abstract: The Scott Polar Research Institute Library has been contributing Antarctic records to the Cold Regions Bibliography Project since 1996. The rationale of the collaboration project is outlined in the context of increasing coverage and reducing duplicate indexing. Although the different requirements of the CRBP and SPRI databases have been seen as a barrier to the sharing of records, a low-tech solution based on word-processing software has been used to adapt records created for CRBP to fit the SPRI format. The in-house advantages to both organizations will be assessed, as well as the implications for future collaboration projects.

It has long been acknowledged that there is considerable overlap of interests between the Antarctic Bibliography produced by the Cold Regions Bibliography Project (CRBP) at the Library of Congress in Washington, and the Antarctic material catalogued by the Library of the Scott Polar Research Institute (SPRI) in Cambridge (Thuronyi & Galpin 1988). Although there is a history of collaboration projects between the two organisations, I will discuss only the specific project in which I was involved, which could be seen as a prototype for other collaborations where one main "host" database does most of the cataloguing, assisted by contributions from other libraries.

After discussions between the two organisations in 1994, it was agreed that SPRI would contribute records from non-English language journals to the Antarctic Bibliography using the CRBP format. The project had advantages for both organisations. Quite apart from practical assistance with time-consuming entries which would bring about cost savings for their sponsors, the Antarctic Bibliography hoped to increase its coverage of non-polar and non-English language journals using some of SPRI's resources. For its part, SPRI gained on-line access to the COLD database at a time when it was not generally accessible via the Internet, and was able to greatly reduce duplicate cataloguing of journals and offprints. Both organisations benefited from being seen to be trying something new.

Although I am the Antarctic Bibliographer at SPRI, I am a linguist rather than a librarian, and this influenced the type of journals which SPRI chose to input. The Scott Polar Research Institute Library has a large collection of Russian language materials, and of more than 15,000 Russian language items in the database, over 1,000 are about

Antarctica. We have catalogued articles on Antarctica and the Southern Ocean from 36 Russian language journals, ranging from *Antarktika* and *Materialy Glyatsiologicheskikh Issledovaniya* to less expected titles such as *Morskoy Flot* and *Priroda*. However, over the last few years some of the relatively high-yield Russian journals have become difficult to obtain. We therefore selected *Russkaya Antarkticheskaya Ekspeditsiya. Informatsionnyy Byulleten'* and *Antarktika* as the basis for SPRI's contribution to the Antarctic Bibliography since both titles have continued to appear fairly regularly. I should point out at this early stage that cataloguing these journals can be quite time consuming, for example, articles in *Informatsionnyy Byulleten'* do not have author abstracts.

Having selected the titles I was to catalogue, my next task was to learn to work with the CRBP database (COLD), which uses the *STAR* system as opposed to the *Muscat* used at SPRI. The differences between database systems have often been seen as a barrier to sharing records. Some of these differences arise from the fact that not all databases have the same purpose, while other differences are of a more mechanical nature.

The SPRI database is basically a catalogue of what is available in the Library. The abstracts are short and indicative, there are no entries for author affiliations, but all entries have a shelf location. When users find an item of interest they can go straight to the journal and read the article in question. COLD on the other hand is a bibliography rather than a catalogue, and as such has different requirements. The abstracts in COLD are longer and substantive, and the author affiliation is given. These are obviously important features when users do not necessarily have immediate access to the full text of an article.

The mechanical differences between the SPRI and COLD databases are more problematic. Their different purposes (catalogue vs. bibliography) mean that the databases require different information, so each has fields that have no equivalent in the other. SPRI and COLD use different field tags for the same information; for example, the author is given at *a in SPRI and at **NAME** in COLD. The systems used for transliterating Russian are different, as are the codes for entering diacritics. Another major difference is that whereas SPRI indexes using UDC numbers (to which the system automatically adds key words) and can add Linnean terms, COLD uses key words selected from a list. Table 1 shows some examples of the various differences.

Table 1. Examples of some differences between SPRI and COLD

SPRI	COLD
Different information	
*loc [shelf location]	AUAFF [author affiliation]
*size [in cm]	REF [number of references]
Different field tags	
*a [author's name]	NAME
*k [indexing terms]	SUBAN
Transliteration and diacritics	
Byulleten'	Biulleten
Teshebayev/Yegor B.	Teshebaev, E.B.
rayon	ra^D&ion
\vAkademik Fedorov\n	^IAkademik Fedorov^R
UDC numbers + keywords	**Key words**
[*for the breeding behaviour of Adelie penguins*]	
598.2 Birds	Aves |b Sphenisciformes |c Behavior
598.45 Penguins [Pygoscelis adeliae]	Aves |b Sphenisciformes |c Breeding
591.5 Animal behaviour	cycles and reproduction
591.551 Animals, breeding behaviour	

The initial intention was that I should add records to the COLD database on-line. However, this did not prove to be a practical proposition, as the Internet connection invariably experienced hiccups. Even a short waiting time half-way through entering an author's name was difficult to cope with, and having the connection timed out sometimes meant that an entire abstract was lost. That method of working was soon abandoned.

Eventually COLD came up with a method for loading a batch of records from a word processor file. My next step was therefore to make a COLD template for the journal articles I would be entering. There are 43 possible fields in COLD, but I was pleased to find that the COLD template for *Antarktika* only requires 31 of these. (I was slightly less pleased when I compared this with the 13 fields required by SPRI for the same journal template!). The disadvantage of batch loading was that the records had to be checked individually at COLD. However, the method worked well enough and over the past two years I have contributed records for both *Antarktika* and the *Informatsionnyy Byulleten'*.

As I mentioned earlier, these particular journal entries can be very time consuming, and I was sometimes spending up to 30 or 40 minutes to complete a record in the COLD format. However, none of this work showed up in the SPRI database. I therefore started

213

to look at ways of converting the records I was producing for COLD into the SPRI format.

Having checked that COLD was not sensitive to the order in which the fields were presented, the first step was to rearrange the COLD field tags to match the SPRI format, and to move to the end of the template the 18 COLD fields that were not needed in SPRI. Once a batch of records had been sent to COLD, these surplus fields were removed from the word processor file with a quick *click, drag, and delete*. This left a record which contained only the information which SPRI wanted, in the order it wanted it.

The next step was to change the field tags from COLD (e.g. NAME) to SPRI (*a) format. This required eight *find & replace* commands, which also covered other minor adjustments such as pagination. Changing the transliteration system and the diacritics took a further eleven *find & replace* commands. (Though these created a number of spelling errors, they were all picked up by the spell checker.) Thus with a block deletion and 19 find & replace commands, a record created in the COLD format could be converted for loading into the SPRI database.

The process was obviously very labour intensive. In fact it took about 40 minutes to carry out all these steps. If the procedure had converted just one record, it would not even have been worth considering. However, using *find & replace all* with 20 records in one word processor file, the time per record is reduced to about 2 minutes, which makes it a more viable proposition.

Even after all these changes, the conversion was still not complete. The remaining alterations involved the punctuation of the author's name, entering UDC numbers and assigning the entry to a main subject category. These had to be done individually and brought the time spent per record to something over 5 minutes. It was felt that the conversion exercise was only a semi-viable option, possible only for time consuming records such as those from *Antarktika* and the *Informatsionnyy Byulleten'* where it represented between 13% and 17% of the time spent creating the original record.

To speed up the conversion process, the next step will be to use macros or bibliographic software to make the necessary changes. However, it will still be quite a lengthy procedure, and though there is software which deals very well with different ways of entering the author's name, changes involving key words, UDC numbers, and main subject headings will still need to be made individually for each record.

Contributing records to the Antarctic Bibliography was, however, only one part of the collaboration project. The two other aims of the project were to reduce duplication of cataloguing effort; and to increase coverage.

SPRI receives 85 journals whose high yield of Antarctic articles has put them on the list of journals which are systematically monitored by COLD. Together they yield an average of 450 Antarctic articles every year, all of which are given full entries in the Antarctic Bibliography. Knowing that they are already taken care of, for the past three years SPRI has not been making catalogue entries for Antarctic articles in these journals, though the journals are available in the Library. A quick calculation shows that by cutting out duplicated cataloguing effort of between 15 and 20 minutes per article, SPRI has saved between 112 and 150 hours per year.

A similar effort has been made to reduce duplication in cataloguing offprints. 74% of Antarctic offprints received at SPRI are given full entries in the Antarctic Bibliography. Though these offprints are entered in the SPRI database they do not get an abstract, and this can save about 10 minutes per offprint. As there are on average 180 items in this category every year, this has resulted in a saving of approximately 30 hours at SPRI.

Thus the time saved at SPRI by reducing duplicated cataloguing effort is approaching 180 hours per year, which is the equivalent of approximately 23 full working days.

Of the 26% of SPRI offprints which did not appear in the Antarctic Bibliography, 10% were outside COLD's interests, many of these relating to the Falkland Islands. However, the remaining 16% had been missed by COLD, and this brings us to the third aim of the collaboration project; increasing coverage of Antarctic material.

A recent study of the Antarctic Bibliography (Hibben, 1997) estimated that whereas coverage of polar journals was almost 100%, between 15%-18% of articles in non-polar English language journals were missed, as were approximately 36% of Antarctic items in non-English language journals. Some of the time saved at SPRI by not duplicating cataloguing effort was used firstly to notify COLD of the items I knew it was missing, and secondly to increase coverage of non-English language material and grey literature.

As journals and offprints came into the SPRI Library I used lists of journals monitored by COLD, supplemented by data on the number of hits and the date of the most recent hit, to identify items likely to be missed. Over a three-year period, SPRI notified COLD of 364 such items, increasing from 79 in 1996 to 118 in 1997 and 167 in the first eight months of 1998.

The number of items we were finding for COLD increased as our own coverage of some less obvious sources of Antarctic material improved. These included grey literature, such as internal reports on sub-Antarctic islands from the New Zealand Department of Conservation, and Australian Geological Survey Organisation reports on marine geology in the Southern Ocean. We also traced a number of publications from the Deutsche Geodätische Kommission that had been missed, as well as several papers published in

French language journals by members of the Laboratoire de Paleobotanique du Mesozoique.

Although these are just a few examples of the kind of Antarctic material we were able to capture, I should add that increasing coverage is not just an aim in itself, but important for the researchers who use our databases, and tend to regard them as even more comprehensive than they could ever claim to be. If duplication of cataloguing effort is seen as less than optimum use of time, how much more so is duplicated research.

To finish off this account of a collaboration project, I would like to point to ways in which collaboration could perhaps be made easier. There are probably between 2,300 and 3,000 articles on Antarctica published in one form or another every year, and though it is not a vast undertaking to capture them all in a bibliographic database, funding for such projects is unlikely to increase. Collaboration is therefore likely to become more rather than less desirable, in spite of the difficulties I have mentioned.

Assuming a similar model were to be started from scratch, i.e., most of the cataloguing is done by a main "host" database, with some assistance from other libraries or institutions, collaboration would, I feel, be easier if:

1. the host database were capable of accommodating a wide variety of record formats with relatively minor editing;

2. the host database could allow both checking and loading of records in batches;

3. the contributing organisation could choose between two methods of record contribution:
    a) using the host database template to create records in the host's own format, or
    b) contributing records created in its own database, for the host to convert.

REFERENCES

Hibben, S.G. 1997. Journal coverage in the Cold Regions Bibliography. In: *Creativity, Lighting the Poles: Collaborative Solutions to Common Problems: Proceedings of the 16th Polar Libraries Colloquy, June 17-22, 1996, Consortium Library, University of Alaska Anchorage and Partial Proceedings of the 6th Northern Libraries Colloquy, July 12-15, 1996, Elmer E. Rasmuson Library, University of Alaska, Fairbanks.* (ed. by J. Braund-Allen & C. Innes-Taylor), pp. 19-22. University of Alaska, Anchorage, AK.

Thuronyi, G.T. & Galpin, V.J. 1988. Cooperation - how much benefit? In: *Twelfth Northern Libraries Colloquy; 5-9 June 1988.* (ed. by A.M. Brennan & M. Andrews), pp. 63-67. World Data Centre A for Glaciology [Snow and Ice], Boulder, CO.