Indexing the proceedings and publications of the Scottish Parliament

Tori Spratt and Shona Skakle

This article describes the processes followed in deciding whether to index the proceedings and publications of the Scottish Parliament. The methods chosen once the decision to index was taken are outlined, along with a discussion of the reality now that the parliament exists and there is material to index. The article concludes by discussing the challenges of organizing and providing subject access to this information and possible future developments.

In October 1998 Janet Seaton and Tori Spratt began work on setting up the Scottish Parliament Information Centre (SPICe). Elections for the Scottish Parliament were to be held on 6 May 1999, and SPICe was to open its doors to Members of the Scottish Parliament (MSPs) on 8 May. During the first two weeks it was necessary to devise staffing structures and advertise the posts, draw up budgets, and develop a service plan. The implementation team had grown to four by the end of the year, and the first of the newly appointed staff arrived in February 1999. How the proceedings and publications of the Scottish Parliament would be indexed was just one of many decisions that had to be taken at an early stage in this process.

The first stage was to identify what the outputs of the parliament were likely to be, and how they would need to be treated to make them accessible to MSPs, their staff, staff of the parliament and the outside world. Although the parliament did not yet exist, it was reasonable to assume that the outputs would, on the whole, be similar to those of Westminster.

It became obvious that just publishing these outputs on the parliament’s website would not provide adequate subject-based access to the material. The only way to provide this would be to index it. Once it was established that subject indexing would be required, the next step was to decide whether to start from scratch or use (possibly with adaptations) a pre-existing system.

Factors to consider included:

- time-scale;
- the cost of developing a local system;
- lack of in-house expertise to develop such a system;
- how the information would be stored and organized;
- whether or not the system used would rely on thesaurus control;
- whether subject indexing was to be automatic or intellectual;
- the need to retrieve information about the UK parliament (Westminster), the Welsh Assembly and the Northern Ireland Assembly.

An early decision was that, whatever route was taken, thesaurus control was vital to maintain the integrity of the data held. This control extends beyond subject access and includes the way that names and organizations are held, thus preventing the problems that can occur where people and organizations are known by more than one name. It also reduces the possibility of inputting errors as records containing typographical errors would not be accepted by the database, and alternative forms of names used would be switched to the preferred form.

The desire for thesaurus control, along with an appreciation of the time and costs that would be involved in developing a local database that met our requirements, meant that the logical decision was to use POLIS (the Parliamentary On Line Indexing Service) to describe and index the proceedings and publications of the Scottish Parliament. The database structure also meant that it should be possible to input and index material with minimal changes.

Once POLIS was chosen, the next decision was whether to employ indexers locally, or pay for a member of staff in Westminster. Although there were likely to be technical difficulties to overcome, SPICe decided to employ indexers locally, feeling strongly that local expertise in POLIS searching and indexing techniques would benefit the whole team. Even at such an early stage in the process it was also obvious that although POLIS was the sensible choice initially, in future a local database might be developed. Such a development would be easier if local indexing expertise were already available.

POLIS (Parliamentary On Line Indexing Service)

POLIS is a sophisticated database that provides name- and subject-indexed references to parliamentary proceedings and publications. It has been in use since 1980 and is available outside parliament, on subscription, as Justis Parliament. Apart from the breadth and volume of material covered, the greatest strength of POLIS from the point of view of SPICe was its thesaurus.

POLIS uses a number of thesauri to control the way that names and subjects are described. These include members, corporate authors, personal author names, organizations,
subject index terms and identifiers. A further strength of the thesaurus is its ability to indicate relationships between terms. For example:

| BT | broader term |
| NT | narrower term |
| RT | related term |
| SN | scope note |
| UF | used for (a synonym) |

The thesaurus is constantly developing and the SPICe indexers will be able to have some input to this development, by suggesting the inclusion of terms that occur first in the proceedings of the Scottish Parliament.

The proceedings and publications of the Welsh Assembly and Northern Ireland Assembly are also being recorded on POLIS. A network of information and IT specialists from the devolved assemblies and Westminster has been in existence since November 1998. In spring 2000 this group was formalized as the Inter Parliamentary Steering Group. It is hoped that this group will assist with the development of cooperative working amongst the assemblies, and between the assemblies and Westminster.

Scottish Parliamentary Profiles database (SPP)

SPP is the SPICe in-house database. Development began in early 1999, although the need for such a database was identified some six months earlier. SPP provides access to a variety of information about MSPs including:

- biographical details and information on parliamentary and government posts held;
- records of electronic voting in the chamber;
- register of interests.

It also provides an enquiry-tracking mechanism for SPICe staff.

As with the proceedings and publications of the parliament, it was felt that thesaurus control would be vital if SPP were to become a valuable resource. We were aware that immense time and effort would be required to develop a local thesaurus from scratch. By the time SPP was being developed, the decision to use POLIS had already been taken. Fortunately, staff in the House of Commons Library were prepared to provide an electronic version of the POLIS thesaurus that could be loaded into SPP.

It was decided at this stage that, as far as possible, all name and subject indexing carried out by SPICe would be based on the POLIS thesaurus. It was felt that the existence of a common indexing language, and thus a common search language, would be beneficial to the SPICe indexers and, more importantly, to the users of our services.

Horizon (library management system)

It was decided to use POLIS index terms as subject headings for the catalogue records to provide consistency with other activities in SPICe. Initially, the POLIS Unit in Westminster sent Word documents containing subject, organization, name and identifier thesauri. These proved somewhat unwieldy to use; without access to the POLIS database itself the cataloguers were unable to find precedent for how to apply the indexing terms. In the summer of 1999, SPICe finally gained electronic access to POLIS and it became easier to understand how the thesaurus worked.

POLIS indexing terms are used to provide subject access to all catalogue records. POLIS indexing is predominantly post-coordinate, which allows for very close indexing and provides flexibility in the retrieval of information from the database. Unfortunately, such a post-coordinate approach is not well suited to the Horizon system (or indeed to many library catalogues). This has been overcome by constructing subject strings from POLIS terms. Work is still progressing with the subject strings but with the arrival of a part-time cataloguer in July 2000, the Bibliographic Services Librarian should have more time to devote to expanding the strings already used.

Scottish Parliament material is indexed on POLIS by the Scottish Parliament indexers. To prevent duplication of effort and maintain consistency across databases, the indexers let the cataloguers know which index terms have been used so that the same terms can be used in subject strings for the catalogue records. In the case of Westminster, Welsh and Northern Irish material, the cataloguers search the POLIS database, copying the POLIS index terms used on records for conversion into subject strings on Horizon. For non-parliament material, POLIS terms are used more freely, although the cataloguers search both Horizon and POLIS for precedent when adding subject headings to such material.

The POLIS thesaurus is used as the authority file for the Horizon catalogue, and is added to on an ad hoc basis. The POLIS thesaurus is particularly strong on organizations and subject headings. However, its personal name thesaurus is much weaker (names are added to the thesaurus only when there is a conflict or when clarification is required). The names thesaurus is therefore not used as the authority file for names in Horizon. Names are added according to the data available on the publication or from other bibliographic sources.

The POLIS thesaurus works well for the type of material the Scottish Parliament receives and it is hoped that users will appreciate the consistency of approach across SPP, POLIS and Horizon.

Bibliography

The bibliography of Scottish Parliament publications will appear every six months and is a joint venture between SPICe and the Stationery Office. To speed up production of
the first issue, an automated index generated by the Stationery Office has been used. POLIS terms will be used in future if a decision is made that there should be intellectual input to the process.

Scottish Parliament Intranet (SPEIR)

SPEIR (The Scottish Parliament Electronic Information Resource) is the main method of delivering information to MSPs, their staff and the staff of the parliament, both directly and as a gateway to the Scottish Parliament website. Although users of SPEIR can access information by navigating around the site in the traditional way, or via the SPEIR search engine, a number of specialist services hosted on SPEIR provide more sophisticated and controlled subject access. The indexing underlying this is on the whole based upon the POLIS thesaurus.

SPICEWeb is a browser interface to the MSP information held on SPP. It includes a flexible search function that allows complex searches to be run. One of its strong features is its use of thesaurus control on most of the data held. Horizon Webpac is a browser interface to the SPICe catalogue and should be available on SPEIR in late 2000.

At present, subject access to SPICe outputs, including research briefings, is provided only at a very general level. The research briefings are also published on the Scottish Parliament website, where the same level of subject access is provided. Once time allows, the briefings will be catalogued on Horizon, allowing subject access via the Horizon Webpac. Although the SPICe catalogue is not directly available on the website, the same level of subject access will be provided.

The Scottish Parliament Website (http://www.scottish.parliament.uk)

One of the aspirations of the Consultative Steering Group was that the Scottish Parliament should be open and accessible. This was taken into account when the website was initially being designed. All parliamentary proceedings and publications are available via the website, which already holds a large amount of information and is constantly growing. Plans for it to become database-driven are well advanced. This should improve access to the information held. For example, MSP biographies, the register of interests and information about MSP voting records are held as, or within, individual web pages. Once the website becomes database-driven it should be possible to import the data structures and subject indexing from SPICEWeb, thus giving external and internal users the same level of access.

Official Report and other parliamentary proceedings

At Westminster, the daily parts of the Official Report (Hansard) are cumulated with corrections to form a fortnightly bound volume and index. The index is generated from POLIS using a specific index field. In the Scottish Parliament the equivalent product will be produced on a monthly basis, and will be a web-based and CD-ROM product. The intention is to structure the material well using a database, and develop a flexible search interface to facilitate retrieval. Indexing will not be from POLIS, as this would require a large amount of development and programming work. Initially, the index is likely to be ‘automatically generated’ from headings used in the Official Report. It should be possible to adopt a similar approach to committee proceedings and reports. Once the website database is developed, it should be possible to apply POLIS-style thesaurus control and subject indexing to all of the proceedings and publications held on the website.

Challenges

The decision to use POLIS for indexing the proceedings and publications of the Scottish Parliament has brought a number of associated challenges. Perhaps the most obvious was not having direct access to POLIS until August 1999. Technical problems still exist, most notably the fragile nature of the connection to the Westminster network and POLIS. A number of intermediaries are required to resolve any problems, which can be frustrating for all involved, and has a detrimental impact on the speed at which material can be added to POLIS and indexed. It has been necessary to match the needs, priorities and standards of the Scottish Parliament, Northern Ireland Assembly and Welsh Assembly to those of Westminster. The challenge of coordinating the development of a large database to do this cannot be overestimated. It can also be difficult to maintain consistency of indexing standards when located at a distance from the other indexers.

There has also been the challenge of trying to adapt a pre-existing system to suit our own proceedings and publications, which, although similar, are not always the same as those in Westminster. In the early stages this was exacerbated by the need to take decisions about the treatment of particular categories of material in advance of (or very soon after) their first appearance.

The quantity of material requiring indexing has been far greater than expected. There has been an increase in the number of written parliamentary questions during the first year of the parliament; committees have sat more frequently than was initially planned; there have been more plenary debates, and a large number of petitions have been lodged with the parliament. To compound the problems, in-built backlogs in indexing this vast amount of material have been caused by a number of factors, including the need to send staff to Westminster for training, the initial lack of access to POLIS, and fragile IT connections.

Future plans

When SPICe was initially examining the best way of indexing Scottish Parliament proceedings and publications it was decided that, as there was a pre-existing database designed for the same purpose, there would be little point in trying to create a new one. Access to POLIS would be needed by SPICe to trace Westminster material, and there was an awareness that each assembly would potentially need access to four separate databases if everyone were to develop their own method of organizing proceedings.
Obviously POLIS will be a much more valuable resource if it continues to hold details of the proceedings and publica-
tions of all the devolved assemblies as well as those of the
UK parliament at Westminster. However, for the devolved
assemblies it is effectively a legacy database, with all of the
attendant problems. The devolved assemblies might be seen
as the ‘junior partners’ in POLIS, which could lead to diffi-
culties in attempting to have any real influence on develop-
ments to the database or even the thesaurus.

At present, all Scottish Parliament records on POLIS are
created manually (through a combination of cutting and
pasting from the website and Word files and some direct
keying). This is a time-consuming process and one that
increases the likelihood of error. The possibility of develop-
ing automatic data capture has been explored, but any
such work would be costly and would have to fit in with
existing POLIS development plans and timetables.

After a year of parliamentary activity, the Scottish Parlia-
ment decided that it would be more time- and cost-effective
to concentrate efforts on developing our own database of
proceedings and publications, rather than waiting for the
developments of POLIS that are needed to allow us to work
efficiently. Subject indexing would still be based on the
POLIS thesaurus. Westminster could then develop a
method of downloading records from this database onto
POLIS.

The decision to develop a local proceedings and publica-
tions database may provide other advantages. At present,
information about MSPs, the SPICe catalogue, and the
proceedings and publications of the parliament must be
accessed via three databases, plus the Scottish Parliament
Website and Intranet. This demands a fairly sophisticated
understanding of how the information is organized.

Although we have attempted to provide coherence and
consistency through the use of a common thesaurus for
indexing the proceedings and publications of the Scottish
Parliament, the current variety of access points required is
undoubtedly confusing for both users and staff. Initial exam-
ination of data structures for the new database design has
shown that it would be desirable to create a more seamless
method of accessing information for our users, eliminating
the need for them to understand how the data have been
structured.

Ideally SPICe would like SPP to become a one-stop shop
for our users and ourselves. MSP information, the catalogue
of SPICe holdings, an index of Scottish Parliament proceed-
ings and publications with links to the full-text documents,
and an enquiry tracking system for the use of SPICe staff
would be held on one database (or a number of linked data-
bases) available via a single interface. Coherence and
consistency would continue to be provided through the use
of name and subject indexing based on the POLIS
thesaurus. Users would have the advantage of not needing
to know exactly what type of material they were looking for,
and which database it was stored in, before beginning a
search. The development of such a database will be complex
and time-consuming, but it is to be hoped that it will provide
an extremely powerful tool for all users of information about
the Scottish Parliament.

Tori Spratt is Resources Manager, Scottish Parliament Information
Centre. Email: Tori.Spratt@scottish.parliament.uk

Shona Skakle is Bibliographic Services Librarian, Scottish Parlia-
ment Information Centre.

---

**Thomas Gray as indexer**

No matter how diverse or distinct from one another the
material of Gray’s readings or his learning in general
seemed to be at first glance, all of that material could be
organized and tabulated clearly within the pages of his
Commonplace Books or would, similarly, find an
appropriate place within the discrete and aesthetically
pleasing columns of his Chronology. The indices [sic]
generally prefaced to such Commonplace Books typically
used the artificial patterns of language and linguistic order
to separate and to codify experience, in much the same
way that the governing structure of the Chronology as a
whole rested on the culturally determined patterns and
(now antiquated) divisions of the Olympiads. . . .

Gray initially attempted to keep his notebooks in the
manner first recommended by the philosopher John Locke
in his writings on education. According to Locke’s method,
subjects were filed in an index prefixed to each volume by,
respectively, first letter and then first vowel. Thus the page
number indicating an entry on Aristotle, for example,
would be filed under ‘A’, and then under a subdivision ‘i’;
an entry on Athenaeus under ‘A’ and then ‘e’, and so on.
Such a system allowed the note-taker to move fluidly
between his daily readings, devoting his attention to any
number of different authors, subjects, or volumes in the
course of the day, while at the same time permitting him to
keep close track of the precise location of his various
notes.

From Robert Mack, *Thomas Gray* (© Yale University Press,
2000), reproduced by permission of the publishers.