ALISS *Quarterly*

Association of Librarians and Information professionals in the Social Sciences

Special issue: Supporting 21st Century Information Needs.

Supporting Practitioners in Health and
Social Care conference papers

Jo Wood, Cafcass; Sue Jardine, SCIE; Margaret Anderson

Cochrane Developmental Psychosocial

and Learning Problems Group.

Supporting systematic reviews

King's College London; London School of

Hygiene & Tropical Medicine.

Supporting Research data users

Training Librarians for Research Data Management Support;

Recent thesaurus developments from the UK Data Service.

Special issue: Supporting 21st Century Information Needs.
Supporting Practitioners in Health and Social Care conference papers.

Editorial

Welcome to the latest edition of ALISS Quarterly. It has been published by ALISS (Association of Librarians and Information Professionals in the Social Sciences).

This issue contains papers from the recent ALISS conference: Supporting Practitioners in Health and Social Care which took place on 11th February 2015 at Senate House, London. The focus of the event was to discuss the information needs of practising social workers, social care and medical professional staff, and offer advice on how library and information professionals might best support them. This involved a consideration of how the needs of practitioners differ from academic library users and what adaptations this involves in library service delivery.

The first paper from Jo Wood, described the process by which she created and developed an in-house library for child protection practitioners in a government department. It provided a fascinating history of the evolution of the Library, covering the development of different services, marketing to users and measurement of impact. The second paper from Sue Jardine provided an introduction to the landscape of health and social care today, highlighting the concerns felt by practitioners and focussing upon the range of free services offered by SCIE which could help support their daily work in the field. The final paper from Margaret Anderson offered a wide ranging introduction to the field of systematic reviews. Using the example of her work as Trials Search Coordinator with the Cochrane Developmental Psychosocial and Learning Problems Group, based at Queen's University Belfast she provided insight into the time span and preparation needed for a systematic review, plus insight into how librarians and information professionals can effectively input to and support the process.

All of the papers are published in this issue and the slides can be viewed on the ALISS website at:

http://alissnet.org.uk/2015/02/13/supporting-practitioners-in-health-and-social-care/

Systematic review literature searching is growing rapidly within the social sciences and the second section of this issue provides information on the range of services offered by two universities: Kings College London, and the London School of Hygiene and Tropical Medicine. Both articles offer a practical perspective on the challenges faced by information professionals in designing and developing supportive library services.

The final section of the issue considers another emerging information area for 21st Library and information professionals: research data. Sarah Jones (Digital Curation Centre) and Mariëtte van Selm (University of Amsterdam Library) provide an overview of 5 research data management courses for librarians which were represented at an International Digital Curation Conference (IDCC) in January 2015. They consider the content of the courses and the type of support information professionals will need in the future to attain greater expertise in this field.

Finally Lorna Balkan from the UK Data Archive describes recent projects to improve their two thesauri: the Humanities and Social Sciences Electronic Thesaurus (HASSET) and the

multilingual European Language Social Science Thesaurus (ELSST) considering how this might aid future data curation and retrieval.

We hope you enjoy the issue.

Keep up to date with our website at http://www.alissnet.org.uk and twitter channel http://twitter.com/aliss_info and by subscribing to our free electronic mailing list LIS_SOCIAL SCIENCE at http://www.jiscmail.ac.uk/lists/LIS-SOCIALSCIENCE.html or consulting our website at: http://www.alissnet.org.uk

Heather Dawson.
ALISS Secretary
h.dawson@lse.ac.uk

Build it and they will come: developing an in-house service for practitioners

Jo Wood, Librarian, Children and Family Court Advisory and Support Service (Cafcass)

This article looks at the development of the Children and Family Court Advisory and Support Service Library, from its inception in 2006 to the present day.

Cafcass was established on 1st April 2001 and brought together the family court services previously provided by the Family Court Welfare Service, the Guardian ad Litem Service and the Children's Division of the Official Solicitor's Officer. Cafcass is a Non-Departmental Public Body (NDPB) accountable to the Ministry of Justice and has a remit to support the 140,000 children and young people who go through the family justice system each year. The organisation employs c.1300 Family Court Advisors (FCAs) who are qualified, experienced social workers, across 43 sites in England. Each FCA holds, on average, 27 cases at any one time. Broadly, cases fall into two categories: public law where children are going through care or adoption proceedings, and private law where parents are separated and are unable to agree on future arrangements for their children.

In 2004, the then Cafcass Research Officer put together a proposal to provide the organisation with a library and information service. The NSPCC was commissioned to conduct a library feasibility study, which concluded that the provision of an information service was both possible and necessary, but that it needed to be provided by an external body with an existing collection of resources. The NSPCC stated that they were unable to provide the service, so Cafcass approached other third sector organisations. In 2006, Cafcass entered into a partnership with Barnardo's. The terms of reference document between the two organisations stated that the partnership would:

*"Enable staff….to access up-to-date published information in a cost effective way
Reduce time spent searching for information
Assist staff to keep up to date with the own area of work
Support staff development and training
Promote knowledge management initiatives"*

In August 2006 I was appointed to the post of Library and Information Services Officer and a month later the Barnardo's-Cafcass partnership started providing a remote library service to Cafcass staff across England. The service was an immediate success, with 170 requests for books, journal articles and subject searches completed in the first month of the service. In the first year, 533 Cafcass staff used the library and I dealt with 1,877 requests in total. The service was extremely well-received and supported by Cafcass staff and the following year the library dealt with over 3,000 requests.

However, in the spring of 2008 the decision was made to close the Barnardo's Library and Cafcass were informed that the partnership would end that August. A couple of senior managers in Cafcass persuaded Barnardo's to keep the library open until the following spring while they worked out how to provide an information service to staff. Between May 2008 and March 2009 my role was split between keeping the existing service running

and investigating and eventually implementing a plan to create a new library service for Cafcass. I received terrific support from the Research Officer and the Knowledge, Learning and Development team in Cafcass during this time.

By March 2009 we had found an office that had space for the new library service. The Cafcass offices in London had very limited space, so we rented a room from the Social Care Institute for Excellence (SCIE), which was based near London Bridge at the time. Cafcass had paid for a small number of books and journals to be added to the Barnardo's Library during the partnership and they were moved to the new location. My employment was transferred over from Barnardo's to Cafcass (I was a Barnardo's employee but my salary was paid by Cafcass during the partnership) and on 27th March 2009 I stood in an empty office with no shelving, no computer (no IT infrastructure at all, in fact), very few printed resources, no Library Management System and no journal or membership subscriptions. I did have a box of brand-new stationery that the Research Officer had bought for me, which included library stamps and date labels for books, but no storage cupboards.

I was given six weeks to re-establish the library service as Cafcass were keen to keep providing the service to the staff who had quickly become reliant on it during the partnership. The Barnardo's library model was still workable and I used that to inform how I built the Cafcass Library. One of the biggest issues was transferring the data from the Barnardo's library catalogue to Cafcass. I decided that the Cafcass Library should use the same LMS provider (OCLC) and the same system (OLIB) to make the transition smoother. We were able to buy the data from Barnardo's so that I didn't have to recreate hundreds of user records and thousands of item records. Starting from a baseline of virtually zero resources enabled me to focus the library budget on items and services that the library really needed. For example, I was able to subscribe to a relatively small number of core journals and set up an inter-library loan account with the British Library to acquire articles from other sources. I also developed good relations with SCIE and the allowed me to both borrow resources from them and provided me with opportunities for continued professional development, which were invaluable as a solo worker separated from my organisation.

In May 2009, once the IT infrastructure was in place and the library had some resources to offer staff, the Cafcass Library opened. An official launch party was held that November. By that time the collection had grown dramatically, from 200 to nearly 6,000 items thanks to some extremely generous benefactors who donated their printed collections to the library. Once the library had been re-established, I had to promote the service again. This involved creating and disseminating a current awareness bulletin, visiting Cafcass offices to meet staff, speaking at staff meetings and conferences, and writing articles for the monthly staff newsletter.

A year later, the service was so busy that I was able to recruit a full-time Library Assistant to support my work and take over some of the front-line library service tasks, including responding to book and journal article requests. In June 2011 the library moved to the same location as Cafcass national office, which was situated in the Department for Education building as they were our sponsoring department. While it was great to finally

be in the same location as my colleagues, the staff were located on the 6th floor of the building and the library collection was situated on the lower ground floor which meant a lot of trips up and down the stairs (or in the lift) to retrieve and shelve resources.

It was during this period that library use really accelerated across Cafcass again after experiencing a dip in 2009-10 during the transition period when the service moved from Barnardo's to Cafcass. In 2012-13 the library dealt with nearly 3,900 requests and in 2013-14 this increased to 4,500 requests. This acceleration can be explained by a number of factors: the library was now embedded in the organisation, the library team was more visible, Cafcass was preparing to be inspected by Ofsted, there was a move towards evidence-based practice, and the library collections had increased support staff. In March 2014 the sponsorship of Cafcass moved over to the Ministry of Justice and National Office moved to new premises in Bloomsbury. The library collections are located on the same floor as the library staff again (no more seven floor trips to retrieve items) and library use has accelerated again as more than 5,500 requests have been dealt with in 2014-15.

Although the library service has changed beyond all recognition since its inception in 2006, the remit is essentially the same as that outlined in the original Barnardo's-Cafcass partnership document. The library now has five core functions:

1. To support the development of evidence-informed practice and to assist practitioners in using research in their work
2. To support the CPD (Continued Professional Development) needs of staff
3. To support staff undertaking academic studies.
4. To support the work of National Office functions, including: the National Improvement Service, Policy, Corporate Services, Communications and HR
5. To support social work students undertaking their placement with Cafcass, and the cohort of Newly-Qualified Social Workers in the organization.

The Cafcass Library now has more than 20,000 items in the collection, composed of books, journal articles, government publications, research reports, dissertations and links to online content. We have institutional subscriptions to a small number of journals and also subscribe to the Medline and PsycInfo data bases via Ovid. Between us, the Library Assistant and I monitor more than 80 journal titles and almost 300 websites and online resources. We are still using the OLIB Library Management System and I catalogue books using the Bliss classification system.

In 2013 we conducted an impact study which aimed to prove our continuing worth in the organization. The results indicated that 99% of respondents described the library as being helpful in assisting them with their casework. Where dissatisfaction was expressed, follow-up telephone interviews revealed that the source of this related primarily to the absence of available research (as in, the research simply didn't exist), rather than with the library service itself. Respondents praised the speed with which requests were dealt (our service level agreement states that we will deal with straightforward requests in less than two working days and subject requests in less than five working days), the efficiency of the library staff and the time saving aspects of being able to access an in-house library. In addition, the service was compared favorably with a number of external information services.

Looking to the future, we are investigating the possibility of providing an e-books service to Cafcass staff and we are also considering whether our current Library Management System meets the needs of the service both now and in the future. However, I am so busy running the service that development is often put on the back-burner while we meet demand. I would like an Assistant Librarian to take relieve me of some of my responsibilities so that I can focus on the development work.

There are many advantages to being able to build a library service from scratch: I have been able to own all of the big decisions and tailor the service to fit the needs of the users, no-one tells me how to do my job, I'm the 'Library and information expert' in Cafcass, and I can make changes to the library without having to go through a lengthy approval process. However, failure is not an option, I constantly have to justify my role, my assistant's role and the library itself to both internal and external stakeholders, we are subject to continued uncertainty in the public sector (and this is likely to increase), I don't have another information professional in the organization to bounce ideas off which makes it quite lonely at times, I have had to learn as I go and accept that progress will not always be linear and it's my fault when things go wrong!

Despite the challenges, my role is hugely rewarding and I honestly believe it's the best job I will ever have in the library and information profession. There are very few librarians that can say that they were given the opportunity to build a library from nothing and make it a success. In addition, I am proud to be part of the 'team behind the team' that supports the Family Court Advisors in Cafcass in making the best decisions for children and young people in the family justice system.

How SCIE supports the information needs of practitioners in health and social care practitioners

Sue Jardine, Information Specialist, SCIE

This article is based on the presentation given at the February 2015 ALISS conference, which provided an overview of the challenges faced by health and social care practitioners, and gave examples of how SCIE supports their information needs.

Statistics project demographic changes which will result in increasing numbers of older people - who have no family to help them or whose partners or spouse are unable to provide informal care, who have long term conditions, complex needs and who have dementia. Increased demand for services will create unsustainable costs.

The government has identified 4 key categories, all of which imply or require a more integrated approach to planning, design and delivery of local health and care services.

1) provide person-centred care; 2) tackle unhealthy life styles; 3) provide support and use technology to avoid inappropriate use of health and social care services, and encourage self-help and provision of information and advice to enable service users to manage their own care resources.

Taken from: http://www.socitm.net/system/files/Redesigning%20health%20%20and%20 social%20care%20summary%20January%202015.pdf

From the point of view of patients and service users, health and care services which are coordinated are more likely to meet their needs. 'People want co-ordination. Not necessarily (organisational) integration. People want care. Where it comes from is secondary'. www.nationalvoices.org.uk

The following diagram illustrates how many health and social care professionals can be involved in peoples' care. http://www.alzheimers.org.uk/
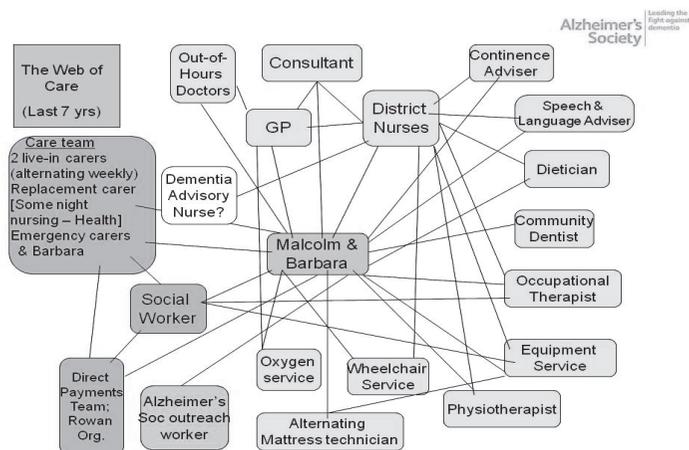
The new statutory framework and responsibilities to implement the categories are contained within the following:

Health and Social Care Act 2012 provides for the most extensive reorganisation of the structure of the National Health Service England to date.
https://www.gov.uk/government/publications/health-and-social-care-act-2012-fact-sheets

The Better Care Fund (BCF) announced at the Spending Round June 2013 is a local single pooled budget to incentivise the NHS and local government to work more closely to provide local services for older and disabled people which enable them to be cared for in the community.
https://www.gov.uk/government/publications/better-care-fund-how-it-will-work-in-2015-to-2016

Care Act 2014 effective from April 2015 is the most significant change in the legal framework, funding and provision of adult social care for over 50 years. It is part of a bigger push to reform social care and health and improve integrated care.
https://www.gov.uk/government/publications/care-act-2014-part-1-factsheets

How SCIE provides support
SCIE gathers and analyses research and evidence about practices and policies of what works. Depending on the strength of the evidence, resources and services are developed to support staff, who plan, commission and deliver services. In the case of the Care Act and Better Care Fund, SCIE is working with a range of organisations to support implementation programmes. The majority of our adult social care programmes of work are commissioned by Department of Health (England), NHS (England) and the National Collaborating Centre for Health and Social Care (NCCSC). http://www.scie.org.uk/about/partnerships.asp

Examples of the of resources SCIE produces
SCIE produces practical and accessible online resources to improve the knowledge, skills and practice of health and social care practitioners.

Guides
Key findings, current legislation and examples of what is working well to guide and inform practice. Guides are designed for online use.

Integrating personal budgets for people with mental health problems
http://www.scie.org.uk/publications/guides/55-integrating-personal-budgets-for-people-with-mental-health-problems/index.asp Integrated personal budgets are new, so the evidence about what works is still emerging.

At a glance
Summarise important messages and practice advice in SCIE guides
69: Adult safeguarding for adults: signs and indicators of abuse, 2015
66: Adult safeguarding for housing staff, 2015 (Link back to SCIE Guide 53, April 2014)
http://www.scie.org.uk/publications/ataglance/index.asp

E-learning
Produced between 2007-2013, materials include audio, video and interactive elements such as quizzes and assessments.

Adult safeguarding resource
http://www.scie.org.uk/publications/elearning/adultsafeguarding/index.asp

Social Care TV
Over 130 films which can also be viewed on YouTube. Used for training/presentations or for personal use on tablets and smart phones.
http://www.scie.org.uk/socialcaretv/index.asp

Databases

Social Care Online
Database of resources on all aspects of social work and social care.
Watch the YouTube introductory video and follow us on Twitter @SCIE_sco
http://www.scie-socialcareonline.org.uk/
Help screens are available at
http://www.scie-socialcareonline.org.uk/Help/Help
They include: Using Social Care Online: an overview
View or download ' powerpoint presentation providing an overview of the key features of Social Care Online and how to use them.

Prevention Library
Commissioned by the Department of Health and launched this earlier this year, this resource highlights emerging practice and research in the provision of prevention services in adult social care.
http://www.scie.org.uk/prevention-library/

Visit www.scie.org.uk to find out more about our products and services and follow SCIE on Twitter @SCIE_socialcare.

Supporting authors of systematic reviews

Margaret Anderson. Trials Search Coordinator with the Cochrane Developmental Psychosocial and Learning Problems Group, based at Queen's University Belfast. m.anderson@qub.ac.uk

Systematic reviews (SRs) are well established in the health and social sciences as a means of summarising and evaluating the best available evidence about a specific question.

Searching for evidence for a systematic review differs from a traditional literature review in several respects. The search is more wide-ranging, often producing large numbers of records. Authors of SRs are also expected to go that "extra mile" to seek out unpublished reports. If necessary, they will contact authors to clarify research methods or request additional data. They then sift through large numbers (perhaps thousands) of records, applying pre-determined criteria in order to select only those studies which are eligible to be included in the final analysis.

One of the key principles of all aspects of systematic reviewing is to minimise bias, so the methods used at each stage must be explicit and capable of being reproduced. Writing a protocol is crucial to achieve this transparency as it establishes in advance parameters of the review question and the methods, and acts as a point of reference throughout the process. The search strategy forms part of the protocol and will typically include a list of sources, an indication of the search terms, and any supplementary search methods (for example, hand searching a set of conference papers which are not indexed elsewhere). As each search is completed, the exact search strategy, search dates and details of each source must be documented so that the searches can be accurately reported in the final review.

The Cochrane Database of Systematic Reviews (CDSR)[1] is highly regarded as a source of systematic reviews on health care. Committing to a Cochrane review is a considerable undertaking, particularly since there is also an obligation is to keep it updated. Before embarking on a review, authors must have their title approved by the editorial base of the most appropriate Cochrane Review Group (CRG). There are currently 53 CRGs worldwide, specialising in different health conditions, or areas of health care. The editorial base of each CRG manages the progress of the review, provides feedback and generally guides the review team through the life cycle of the review. Most CRGs have an information specialist (Trials Search Coordinator) who provides a range of support from devising a strategy to running some or all the searches.

Cochrane methods are highly prescriptive and, in addition to the guidance provided by the Cochrane Handbook[2], there are sets of standards[3,4] which specify how reviews should be conducted and reported. Standards are classified as either "mandatory" or "highly desirable".

1   *Cochrane Database of Systematic Reviews, part of The Cochrane Library. www.thecochranelibrary.com*
2   *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org*
3   *Methodological standards for the conduct of new Cochrane reviews (version 2.3 12 December 2012) http://editorial-unit.cochrane.org/mecir*
4   *Methodological standards for the reporting of Cochrane intervention reviews (version 1.1 17 December 2012) http:// editorial-unit.cochrane.org/mecir*

As Trials Search Co-ordinator (TSC) for the Cochrane Developmental, Psychosocial and Learning Problems Group (CDPLPG), it is my job to check that the search sections of the review comply with the standards which relate to searching for studies. My first official contact with authors is when they submit the first draft of the protocol, though they might already have been in touch to ask for guidance. I check they have included the mandatory set of databases which must be searched for every review, but also expect to see subject specific databases relevant to the review, as well as grey literature sources e.g. theses and conference abstracts. It is not unknown for lists of databases to be copied from previously published reviews, so I now remember to check with authors whether or not they have access to everything on their list!

The published protocol will include a search strategy written for a key database such as Ovid MEDLINE. The strategies for intervention reviews are structured using the main concepts of the review (e.g. population AND intervention). As this will later be adapted for other databases, it is important to make sure there are adequate search terms to capture relevant records. Our diverse scope makes this the most challenging but interesting part of my job, as our reviews focus on a wide range of conditions (e.g. autistic spectrum disorders, malnutrition, speech disorders, children who are maltreated). The protocol is an obvious source of search terms, but I frequently seek clarification from the authors, particularly if the intervention is complex. I also rely on their subject expertise to suggest appropriate synonyms, related or regional terms. Devising a strategy can sometimes involve a number of small test searches being sent to the authors so that they can compare the records retrieved by different combinations of search terms. It is easiest when the review team have help from a local librarian, as it is challenging to explain MeSH terms and the finer points of search syntax by email!

To keep the evidence in the review as current as possible, searches have a "shelf life" of 12 months. If the review is not published within this time, the searches need to be updated. Ideally, therefore, there should be as little delay as possible between running the searches and beginning the screening process. Managing the searches and managing the bibliographic records needs to be carefully thought out in advance. Agreement needs to be reached about who is searching which database, who is managing the records and coordinating the screening process, and which reference manager software will be used. Screening (or sifting) the records needs to be methodical as the completed review will have a study flow diagram to illustrate the number of records which were identified and how many records were eliminated at each stage of the process.

When authors run the searches, I remind them to refer to the relevant standard so that they record not just the name of the database but also the provider, not just the date of search but the date range or issue number of the database. The exact search strategy for each resource should also be reported in the final review. Eagle eyed copy editors will also be checking for these details just before publication, so accurate record keeping is essential!

I prefer to collate the search strategies in one document as I complete each search; hunting though various files and databases accounts for saved searches at a later stage is arduous. The act of pasting the searches to another location can sometimes highlights an error which I can then rectify immediately.

Of course, Cochrane is not the only publisher of systematic reviews. Other organisations, such as The Campbell Collaboration and The Evidence for Policy and Practice Information and Co-ordinating Centre (*EPPI*-Centre) specialise in producing systematic reviews for the social sciences. NGOs and government departments may commission reviews to inform policy. SRs are also published in journals. Judging by the anecdotal evidence of social science librarians, systematic reviews are also becoming a popular choice for theses and research projects.

SRs are underpinned by effective searching of a comprehensive range of databases. This is a daunting task for any prospective review author, even those with previous searching experience, so it is hardly surprising that social science librarians and health librarians are increasingly being approached for guidance and support.

In conclusion, information professionals can provide input to the systematic review process at a number of stages:

1. Awareness of sources – subject librarians are best placed to advise on database content and coverage, and the limitations of some resources for systematic searches (e.g. those databases which limit the number of records which can be downloaded at one time or which have limited search options)
2. Searching expertise – understanding of controlled vocabulary, ability to exploit features such as truncation and proximity operators, ability to adapt searches for different databases and platforms.
3. Knowledge of database features – how to store searches, download strategies and download records in suitable formats
4. Reporting the process – documenting searches, citing references
5. Monitoring impact – advice about altmetrics and how to setting up more traditional citation alerts of a published review

Systematic review training for library users at King's College London: the past, present and future

Marlene Blackstock, Library Learning & Teaching Manager & Sonya Lipczynska, Library Liaison Manager, King's College London

Systematic reviews underpin much of healthcare practice and research, and one of the most reliable sources of evidence (Clarke 2011). Creating a literature search strategy as part of this process is a more complex affair than a straightforward literature review: project participants will need to consider subject heading searches and free-text searches, the use of field suffixes, floating sub-headings and specially created filters to locate a very accurate, detailed and specific set of data. For many setting out on this project, this will be the first time that they have had to create such specific search queries, or work with such large amounts of literature.

We have a large healthcare research community at King's College London, who are supported by us a designated Training & Skills team in developing and maintaining their searching and referencing skills. In 2011 we found that individual training requests for advanced search skills were increasing. Research staff and students were performing scoping searches and developing complex searches as they began long-term systematic review projects and we decided to create a formal class as part of our 'iSkills' library training portfolio. This would fulfil an identified need among our users, as well as enabling our team become more familiar with the process of systematic reviews and the information skills needed to begin a project of this nature.

We began by pooling our knowledge of our user base and the questions they were asking to create a pilot session. This was a single session of 3 hours, which included the following topics:

- Initial process of searching for materials
- Suggested databases (e.g. Medline, Embase, PsycINFO, Web of Science, Scopus, CINAHL, Cochrane library)
- Advanced techniques (field searching, floating subheadings, filters, truncation tools)
- Searching grey literature
- Contacting authors
- Tools for evaluating search strategies
- Tools for documenting material found
- Tools for critically evaluating material found
- Advice on how to store material and when/if to stop looking

Feedback on the session was largely positive although it also highlighted a particular obstacle we had also observed during the session. We had assumed a level of knowledge of literature searching among the researchers who participated (upon reflection this was based on our own level of knowledge). However, it became clear as the session progressed that their experience in this area was more mixed. As we had pitched the session for those who were comfortable with literature searching techniques, we became

somewhat delayed in delivering our planned programme in order to stop and go over some of the fundamentals of literature searching.

We held a debriefing session afterwards and decided that we should split the session into two parts, which we would try to run within a week of each other. Enhancing Skills for Systematic Review part one would focus on the fundamentals of literature searching, with one eye on the overall systematic review aim. This session covered elements such as MeSH headings v. freetext searching, using truncation tools, Boolean operators, constructing a methodical search and how to save and locate those searches. We also covered the importance of using bibliographic software to save and evaluate data, and looked forward to the topics which would be covered in session two.

This initial session was made compulsory for those users wishing to attend Enhancing Skills for Systematic Review part two – in this way, we were able to assume a level of knowledge and move on to more advanced search techniques (e.g. using floating sub-headings in Medline) without needing to coach the participants through the initial stages of setting up a search. Participants who already had a good level of literature searching knowledge were required to contact us to discuss their prior training, and warned that Part Two would not cover basic literature searching. The second session covered how to use filters to locate specific material (e.g. the Cochrane filters for locating randomised controlled trials), floating sub-headings, providing a toolkit of resources including checklists and flow diagrams to evaluate, document and manage data. We also included a comprehensive section on searching grey literature and discussed what should be included in a write-up (the Cochrane library completed systematic reviews are invaluable in giving participants an idea of what to aim for). We also talked about the amount of data required – this is a difficult question to which the only answer can be, 'how long is a piece of string?', but we recommended that participants used their judgement and that of their colleagues to decide when to stop searching. We also made sure to reiterate that searches should be re-done before submission to minimise any nasty surprises which might emerge after a year of research (e.g. a Cochrane review being withdrawn). This two part structure worked very well for us, not least because we were also able to use the Part One session to teach users who were looking for a course on the fundamentals of literature searching.

A particular challenge which has emerged recently is the rise in the number of healthcare students being set mini-systematic reviews as assignments. This is a difficult issue to manage, largely because the majority of these assignments are not systematic reviews in the pure sense but literature reviews being done in a systematic manner. There was some work to be done around liaising with academic staff and discussing the wording and description of these assignments, and this also raised questions about how we should support these students when our training rooms had limited space. The initial answer has been to include evidence-based practice elements into library classes already embedded in the students' curriculum, or to negotiate new embedded classes which tie in with that particular piece of work. Further, we also witnessed a growing number of social science researchers enrolling on the course which meant we needed to research and include some additional information on qualitative research filters.

These sessions have proved extremely popular with long waiting lists. Recent team changes mean that we are even now reviewing our provision in this area. An e-learning module has been developed to complement the training, which includes the slides and additional materials and tools which will be referred to during the sessions. It is projected that the original part one session will merge with a new and popular Literature Review workshops for postgraduate students designed to explore literature searching skills. Part two will become a single systematic review session aimed at researchers who have literature search skills and who are involved in genuine systematic review projects.

At present, the Graduate School and the Biomedicine Research Centre at King's also deliver systematic review lectures for staff and postgraduate research students. It is anticipated that a benchmarking exercise will be conducted to avoid duplication and to tailor systematic review courses taught by library staff so that they are more focused and provide even more hands-on training and interactivity.

Bibliography

Clarke, J. (2011) What is a systematic review? *Evidence Based Nursing*, 14(3), pp. 64.

Further Reading

Anon, CASP Tools & Checklists. Available at: http://www.casp-uk.net/#!casp-tools-checklists/c18f8

Anon, CONSORT - checklist and flow diagram. Available at: http://www.consort-statement.org/

Anon, CRD guidance for undertaking systematic reviews in health care. Available at: http://www.york.ac.uk/inst/crd/index_guidance.htm

Anon, ClinicalTrials.gov. Available at: https://clinicaltrials.gov/

Anon, Learning from Research: Systematic reviews for informing policy decisions. Available at: http://www.alliance4usefulevidence.org/publication/learning-from-research/

Anon, Music interventions for improving psychological and physical outcomes in cancer patients - Cochrane Database of Systematic Reviews - Bradt - Wiley Online Library. Available at: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006911.pub2/pdf

Anon, OpenDOAR - Directory of Open Access Repositories. Available at: http://www.opendoar.org/

Anon, OpenGrey. Available at: http://www.opengrey.eu/

Anon, PRISMA - checklist and flow diagram. Available at: http://prisma-statement.org/

Anon, PROSPERO – The international prospective register for systematic review protocols. Available at: www.crd.york.ac.uk/PROSPERO/

Anon, The DART - Europe E-theses Portal. Available at: http://www.dart-europe.eu/basic-search.php

Glasziou P., 2001. *Systematic reviews in health care: a practical guide*, Cambridge: Cambridge University Press. Available at: http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=74376

Goldacre, B., 2012 What doctors don't know about the drugs they prescribe. Available at: http://www.ted.com/talks/ben_goldacre_what_doctors_don't_know_about_the_drugs_they_perscribe

Goldet G. & Howick J., 2013. Understanding GRADE: an introduction. *Journal of Evidence-Based Medicine*, 6(1), pp.50–54.

Gough D., Oliver S. & Thomas J., 2012. *An introduction to systematic reviews*, London: SAGE.

Higgins D.J.P.T. & Green S., 2011. Cochrane Handbook for Systematic Reviews of Interventions. , Wiley Cochrane Series. Available at: http://handbook.cochrane.org/

Howland, R. H. 2011. Agomelatine: an example of publication bias. *Journal of Psychosocial Nursing and Mental Health Services*. 49(9): p4

InterTasc Information Specialists' Subgroup. ISSG Search filters. Available at: https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home/search-filters-by-design

Kings College London Library resources, 2015. Systematic Review Guidance for Researchers. Available at: http://libguides.kcl.ac.uk/medicaldatabases

McCrae, Niall, Blackstock, M, Purssell Ed., 2015. Eligibility Criteria in Systematic Reviews: A methodological review. *International Journal of Nursing Studies*. In press corrected proof. Science Direct. Available at: http://dx.doi.org/10.1016/j.ijnurstu.2015.02.002

Ovid Medline Support Center, 2014. List of Medline subheadings. Available at: http://ovidsupport.custhelp.com/app/answers/detail/a_id/2078/~/list-of-medline-subheadings-and-their-codes Wolter-Kluwer Health

Library support for systematic reviews at the London School of Hygiene & Tropical Medicine

Jane Falconer, Information Services Librarian. jane.falconer@lshtm.ac.uk

Introduction

Systematic reviews are well established in evidence-based medicine, where experimental results are used to support decision making. The methodology is gradually spreading to other subject areas, including the social sciences, with the support of organisations such as the Campbell Collaboration[1] and the EPPI Centre[2]. This paper describes the support provided to systematic reviewers at the London School of Hygiene & Tropical Medicine by Library & Archives Service staff.

A systematic review is a specific type of literature review. A clear question is set before comprehensive and transparent methods are used to identify, select and evaluate all of the relevant research. Data is then collected and analysed from each included study to provide unbiased summaries of the results. In this way, reliable evidence-based reviews can be produced and readers can have confidence in the conclusions drawn.

The PRISMA Statement (Moher et al., 2009) has been developed to make sure that all relevant methodological information is included in the published review, and can also be used to guide reviewers through the systematic review process. It is useful when helping reviewers as it provides minimum standards for each stage of the review process, including the literature searching and study selection process.

Training and support provided by the Library

Support is provided for systematic reviewers at the London School of Hygiene & Tropical Medicine by the Information Services Team in the Library & Archives Service, which I lead. We are responsible for information literacy teaching and support to both staff and students, and for the inter-library loans service.

Most of our PhD and Doctor of Public Health students begin their research with a systematic review of their topic. Six hour classes on literature searching, scheduled over two half days are arranged via the Research Degree Transferrable Skills Programme. Most research degree students also make an appointment to meet a member of the team to discuss their individual research topic.

For academic staff, classes on literature searching for systematic reviews are run three times per year via our staff Talent and Educational Development programme. Staff can also arrange an individual appointment with a member of the team.

We also offer support in sourcing the full text of items. We offer all staff and students unlimited inter-library loan requests, although we also encourage reviewers to take advantage of the resources available from other local libraries.

If a research team does not contain the expertise required, I can also be sub-contracted

---

1   http://www.campbellcollaboration.org.
2   http://eppi.ioe.ac.uk.

by research groups to do the search, sourcing full-text and study selection for them. This service has only been offered since the start of 2014 and so far I have worked on five systematic reviews or similar studies.

Common problems encountered by systematic reviewers

We find that reviewers often underestimate how long the literature search and study selection takes. Often when they come to us for help, they are trying to find short-cuts in the process. Obviously a level of pragmatism has to be employed as nobody has unlimited time to complete their review, but the minimum standards as set out in the PRISMA Statement (Moher et al., 2009) usually require several days to complete.

Academic staff often think they have the literature search skills required, however a recent study has shown that librarian co-authors have a large positive impact on the quality of the search strategy used in systematic reviews (Rethlefsen et al., 2015). My own experience of evaluating systematic reviews shows that the reporting of the literature search, even in reviews published in good quality journals, is often below the expected standard. The PRISMA Statement has helped to raise standards as well as provide us with evidence to use when providing feedback on search strategies.

New reviewers may need help in running a comprehensive search, we often provide support in choosing keywords, running subject heading searches and transferring searches from one database to another. Many systematic review literature searches can be over 100 lines long and include hundreds of terms. Many reviewers find this a daunting task, particularly when the search has to be run in several sources. Once the searching is completed, many reviewers require help with, de-duplicating results, sourcing the full text and keeping track of the study selection process. It is not unusual for reviewers to have to screen several thousand studies and the reasons for inclusion and exclusion of each one must be recorded.

Challenges and opportunities

Supporting systematic reviewers requires expert skills in literature searching. Running a comprehensive search usually requires using all of the advanced search options available in each database and having an in-depth understanding of how each database is constructed and the information it contains. We often receive queries from reviewers asking us why their search did not retrieve the items they were expecting and this has often resulted in a close analysis of several hundred search terms across several lines of search.

Working with research groups on their systematic reviews has provided the library with many advantages. It has raised the professional profile of the library staff and most of our appointments now come through recommendations. It has also reconnected many academic staff with the Library. Few of our academic staff come into the library itself as they can access the information they require from their desktop, and so we were in danger of losing that personal link. Providing systematic review support has built personal relationships with staff at all stages of their academic career and helped to ensure the service as a whole continues to meet their needs.

I have found that working as part of a research team on a review has improved the

teaching and support we provide to others. We can use my experiences to understand theirs and view the process from a practical as well as a theoretical standpoint. We also provide far more practical support on managing the search process and keeping track of individual studies, previously our support was centred on creating and running the literature search.

Conclusion

Systematic reviews are an important study methodology within global and public health and the Library at the London School of Hygiene & Tropical Medicine has evolved its services to support them. More than many other methodologies, there are clear areas where professional librarian support can and should be offered to researchers. Often researchers do not have the literature search and information management skills required to conduct a high quality review which meets the minimum PRISMA Statement standards.

Providing support for systematic reviewers has improved library staff skills in using our databases, allowing them to provide practical advice to all our library users. It has also raised the profile of the Library within the institution and improved our relationships with otherwise hard to reach academic staff. However, library staff with an excellent knowledge of each database, the search options available and the effectiveness of these is required. Therefore training and mentoring should be offered to any librarian interested in supporting systematic reviewers.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & The PRISMA Group 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6, e1000097.

Rethlefsen, M. L., Farell, A. M., Osterhaus Trzasko, L. C. & Brigham, T. J. 2015. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *Journal of Clinical Epidemiology*.

Training Librarians for Research Data Management Support

Sarah Jones (Digital Curation Centre) and Mariëtte van Selm (University of Amsterdam Library).
Corresponding author: selm@uva.nl

At the International Digital Curation Conference (IDCC) last January a full day workshop showcased five training courses on research data management (RDM) for liaison librarians and other support staff. By sharing training materials and lessons learned, the speakers hoped to enable participants to run their own courses more easily. The workshop proved to be a valuable opportunity to share experiences and get inspiration. Attendees came from over 15 different countries, including Australia, the United States, Denmark, Estonia and Japan, so it was a truly international discussion.

The role of libraries was a key discussion point, and concerns about providing support on research data management were common across the range of countries. Many of the participants were just getting started so this posed a new area of work for them. Some were daunted by the prospect that they not only needed to learn themselves, but then also had to train colleagues and support researchers. Several of the training courses presented at the workshop demonstrate how relevant librarians existing skills are to allay some of these fears. There is much that can be built on.

Courses

The following five courses were presented at the event. The content from each course is available for reuse and a comparison of the different module in each is available in the table below.

Essentials 4 Data Support	RDMRose	DIY Research Data ToolKit	supportDM	RDM Support
I - Definitions II - Planning Phase III - Research Phase [collecting, storing & describing data] IV - User Phase [archiving, publishing, citing data] V - Legislation & Policy [includes licensing] VI - Data Support [roles, case studies]	1. Introduction, RDM and the role of LIS 2. Nature of research and need for RDM [policy & practice] 3. Digital curation lifecycle [including DMPs] 4. Key institutions and projects in RDM [support landscape] 5. What is data? 6. Managing data [storage, description, citation] 7. Case studies of research projects 8. Institutional case study [library role in RDM]	1. Pre-training [intro & research data explained] 2. Data management planning 3. Organising & documenting data 4. Data storage & security 5. Ethics & copyright 6. Data sharing 7. Independent study assignment [Data Curation Profiles interview]	1. About research data management 2. Guidance and support for researchers 3. Data management planning 4. What data to keep and why 5. Cataloguing data 6. Sharing data	1. Introduction to research, data and RDM 2. Data Management Planning 3. Data storage & retention 4. Data organization & description 5. Rules on research data 6. Data sharing & reuse 7. Data & services [throughout the training: case study]

Essentials 4 Data Support is an introductory course for people who (are going to) support researchers in data management. The course is divided into six chapters and can be taken online or as a full course, which is a combination of e-learning and two full day meetings. Essentials 4 Data Support is provided by Research Data Netherlands, an alliance of three Dutch data archives. It replaces Data Intelligence 4 Librarians, the data management course run by DANS and 3TU data centre.

Course materials: http://datasupport.researchdata.nl/en

RDMRose is a set of learning materials on research data management for information professionals that can be used for taught courses and continuing professional development. It consists of eight sessions and comes with a self-evaluation tool to help those with some prior RDM knowledge to select the most important sections to cover. The curriculum is research-led and inquiry based, encouraging reflection on how issues relate to the librarian's own role amidst other professional services. RDMRose was developed by the iSchool at the University of Sheffield, and was trialled with library staff at the Universities of Leeds, Sheffield and York.

Course materials: http://rdmrose.group.shef.ac.uk/?page_id=10

The Do-It-Yourself Research Data Management Training Kit for Librarians builds on the Research Data MANTRA online course (originally developed for PhD students), adding discussion questions and exercises targeted at academic service librarians. The kit is divided into five sessions and contains a mixture of MANTRA units and exercises from the UK Data Services' *Managing and sharing data* training resources. Pre-training promotional slides and evaluation forms are also provided. MANTRA and the DIY Kit are maintained by EDINA and the University of Edinburgh Data Library.

Course materials: http://datalib.edina.ac.uk/mantra/libtraining.html

supportDM is a blended learning course for people in RDM support roles, consisting of six modules. The online modules were, like Research Data MANTRA, developed in Xerte and can be imported in virtual learning environments. They are accompanied by presentations, exercises, tasks and notes on using the material. The course was developed by Library Services at the University of East London.

Course materials: http://www.uel.ac.uk/trad/outputs/resources

RDM Support was developed for the liaison librarians of the University of Amsterdam Library. The learning materials for this seven session taught course consist of presentations, exercises and video transcripts. Almost all videos used are from Research Data MANTRA. A questionnaire is provided that can be used to gauge prior knowledge before the training and assess the impact of the training afterwards. This questionnaire is modelled on the self-evaluation tool from RDMRose.

Course materials: http://dx.doi.org/10.6084/m9.figshare.1285313

Common themes
Two themes emerged strongly in discussion of the courses: namely the need to view

things from a researcher perspective and the need to build confidence amongst librarians.

One course that has evolved significantly is that offered by Research Data Netherlands. The course was entirely reshaped to focus more on a researcher's perspective: originally Data Intelligence 4 Librarians addressed the skills needed by those who support researchers, whereas Essentials 4 Data Support is framed around researchers' practices and their lifecycle. Participants in the RDMRose course similarly found the DCC Lifecycle Model a bit too curation specific for addressing RDM – a different perspective was needed. A number of exercises were devised to encourage participants to get out of the library and talk to researchers. Others took care to address their colleagues as researchers when delivering training so they could empathise with users. There was consensus across the courses that understanding researchers' workflows and needs was critical to deliver appropriate support.

The need to build confidence amongst librarians that they are equipped to support RDM was another common theme. The *DIY Training Kit* in particular plays close attention to this. Sessions focus on discussion of online modules that participants had considered in advance. By keeping presentations to a minimum, the course designers hoped to avoid the perception that there is an expert on each topic who should be consulted. Discussions served to show that every participant had something to contribute and could draw extensively on their existing skills. Each of the five courses presented included exercises around interviewing researchers, developing guidance websites or responding to fictional enquiries, precisely to reassure participants that they could engage and handle such requests if called on. In light of this core skillset, the RDM Support course focuses explicitly on sharing RDM-specific knowledge so participants are aware of the different policies, legislation and practices pertaining to research data.

Selecting content to reuse

The IDCC workshop included an exercise on prioritising topics to be covered in courses. When discussing which elements to include, participants found it hard to select. There are lots of topics that *could* be included and all were felt to be relevant. The choice largely came down to the target audience. For researchers, delegates focused on topics that emphasised benefits to engage their interest such as citation and impact. Data management planning and policy requirements were selected by most groups too. Discussion picked up on the need for subject-specific examples to set general principles in practice and a suggestion was made to arrange prominent researchers to talk as they would make a more persuasive case for RDM.

Upcoming work

Discussion from the event at IDCC is informing Research Data Management workshops at the LIBER conference in June 2015[1]. The first half-day workshop will address data policy and how RDM is a new leadership role for libraries, while the second will focus on

---

1   For details of the "Libraries and research data: Towards a new leadership role" workshops at the LIBER conference, see: http://www.liber2015.org.uk/event/libraries-and-research-data-towards-a-new-leadership-role-part-1/?instance_id=112

providing support via training and data management plans. The FOSTER project[2] is also collating existing training materials that promote RDM and open science. These materials are being used to develop a number of e-learning and blended learning courses. There is a wealth of material to draw from and much of the content is made available under CC-BY licenses to provide opportunities for remixing and reuse. The clear message from our final panel at IDCC was the earlier you start the better: don't use a lack of infrastructure as an excuse, make use of what has already been shared and get started!

The presentations from the workshop can be found at http://www.dcc.ac.uk/events/workshops/comparing-notes.

---

2   https://www.fosteropenscience.eu/project

Recent thesaurus developments from the UK Data Service

Lorna Balkan, UK Data Archive, University of Essex

Introduction

The data landscape is changing, in particular in the context of the Semantic Web and Big Data. These changes bring new challenges for data curation in data archives and beyond. Thesauri, in particular, are finding new uses and applications. This paper describes recent work that the UK Data Service has undertaken in thesaurus development to meet some of these new challenges. The thesauri in question are the two that the Service currently manages – the monolingual Humanities and Social Sciences Electronic Thesaurus (HASSET)[1] and the multilingual European Language Social Science Thesaurus (ELSST)[2]. The work has been funded by two separate projects based in the UK Data Archive at the University of Essex: the Jisc-funded SKOS-HASSET[3] which ran from June 2012 to 31 January 2013; and the ESRC-funded CESSDA-ELSST[4] project which runs from October 2012 to September 2017.

Thesauri – traditional uses

Thesauri have long been recognised as useful indexing and searching aids in data archives. They support information retrieval in a number of ways. Firstly, as a controlled vocabulary, they provide a common indexing language of descriptors, to which synonyms or near-synonyms can be mapped for the purposes of retrieval. Secondly, their hierarchical arrangement of terms into Broader Terms (BTs) and Narrower Terms (NTs) can be exploited to either expand or narrow a search as required. Thirdly, unlike much uncontrolled language where words can have more than one meaning, thesaurus terms are restricted to one meaning, conveyed either by their place in the hierarchy, or sometimes via a scope note. Thus the use of thesauri supports very precise searching at different levels of granularity.

It is in this traditional role that thesauri have been employed in the UK Data Service. The Service holds the largest collection of social science data in the UK, with over 6,000 datasets of quantitative and qualitative data, and has developed its own subject thesaurus, HASSET. Originally based on the UNESCO thesaurus, IP in HASSET was granted to the University of Essex in the 1990s and it has been developed in-house since then. The Service uses it to index and search its data collection.

The Service also manages the related multilingual thesaurus, ELSST. ELSST is used for cross-lingual information retrieval in the Consortium of European Social Science Data Archives (CESSDA) data portal[5]. ELSST has been funded over the years by a series of EU projects, including LIMBER[6] and MADIERA[7], and most recently by the ESRC. It is currently available in 12 languages (Czech, Danish, English, Finnish, French, German,

---

1 http://hasset.ukdataservice.ac.uk/
2 http://elsst.ukdataservice.ac.uk/
3 http://www.data-archive.ac.uk/about/projects/skos-hasset
4 http://ukdataservice.ac.uk/about-us/projects/cessda-elsst/details.aspx
5 http://www.cessda.net/catalogue/
6 http://www.data-archive.ac.uk/about/projects/limber
7 http://www.data-archive.ac.uk/about/projects/madiera

Greek, Lithuanian, Norwegian, Romanian, Spanish, and Swedish). ELSST was originally developed from HASSET, and English remains the source language.

Both HASSET and ELSST are available to outside organisations on completion of a licence.

Thesauri - new directions

Until recently, vocabularies, including thesauri and other Knowledge Organization Systems (KOSs) such as ontologies, taxonomies, and classification systems, were considered to be tied tightly to particular domains and applications (Dunsire et al., 2012). Increasingly, there has been a shift towards vocabulary integration and reuse, driven largely by the Semantic Web and its related technologies. The Semantic Web is an extension to the existing web, originally initiated by Tim Berners Lee (Berners-Lee et al., 2001) with the purpose of making the semantics of information and services available on the web understandable not only to humans but also to machines.

At the heart of the Semantic Web are Linked Data[8], enabled by technologies such as the Resource Description Framework (RDF)[9] and other data modelling initiatives, and Uniform Resource Identifiers (URIs). URIs are a generic means to identify entities or concepts in the world, thus allowing anyone to link or refer to them. Vocabularies such as thesauri are used to define concepts and the relationships between them (WC3: Vocabularies, 2015) and thus underpin the Semantic Web. Vocabularies can also be linked to other vocabularies thus allowing users to search a repository indexed with one vocabulary using index terms from another (Méndez and Greenberg , 2012).

Simple Knowledge Organization System (SKOS)[10] provides a standard way of representing KOSs using RDF, and is thus compatible with the Semantic Web. SKOS enables KOS interoperability, data sharing, linking and data merging. Many prominent thesauri are now available in SKOS format (see for example W3C: SKOS/Datasets).

Vocabularies, particularly in standard formats and as Linked Data, are finding new applications beyond data integration, including knowledge management applications (see for example WC3: Vocabularies[11] for a discussion). Vocabularies range in complexity from formal ontologies, to loose collections of terms. The complexity of vocabulary needed depends on the application.

Another recent change to the data landscape is the increasing availability of Big Data. The term Big Data has no fixed definition, but in the social sciences is used to refer not just to very large datasets (possibly as the result of data linkage), but to data that, unlike traditional research data, have not been designed for research but are the result of administrative or commercial transactions. The term is also used to cover social media data (see for example the ESRC Big Data Network[12]). Unlike traditional research data, many of these data are either unstructured or semi-structured. This poses challenges for indexing and retrieval. Solutions to annotating Big Data include automatic or semi-

---

8   http://linkeddata.org/
9   http://www.w3.org/RDF/
10   http://www.w3.org/2004/02/skos/
11   http://www.w3.org/standards/semanticweb/ontology
12   http://www.esrc.ac.uk/research/major-investments/Big-Data/

automatic indexing linked to thesauri and other controlled vocabularies (see for example the Semantic Annotation and Mark-Up for Enhancing Lexical Searches (SAMUELS)[13] project), and collaborative indexing, where users collaborate in the indexing of data (see for example Voss, 2007).

Below we discuss how the SKOS-HASSET and CESSDA-ELSST projects have enabled the UK Data Archive to develop and apply its thesauri for the benefit of the UK Data Service in response to these new challenges.

SKOS-HASSET project

As its name suggests, one of the key goals of the SKOS-HASSET project was to convert HASSET to SKOS format. A secondary aim of the SKOS-HASSET project was to apply HASSET to automatic indexing of the Archive data collection.

Converting HASSET to SKOS

The main objective of SKOS is to enable the easy publication of KOSs for the Semantic Web. The SKOS version of HASSET is implemented via BrightStarDB[14] for the triple stores, and each concept is associated with a Globally Unique Identifier (GUID). The thesaurus is published via Pubby[15], which provides a browseable, meaningful view of the thesaurus (Bell, 2012). The SKOS version of HASSET has proved popular with researchers and web developers.

Automatic indexing

The Archive used the Keyphrase Extraction Algorithm (KEA), a text mining and a machine learning tool, to automatically index four different types of data collections in the UK Data Service: (i) the catalogue records; (ii) the Survey Question Bank (SQB) questionnaires; (iii) the Nesstar bank of variables/questions; and (iv) the case studies and support guides. KEA builds a classifier model using training documents with known keywords which is then applied to help assign keywords to new documents. The results were compared to those produced by the human indexer (the 'gold standard') using manual and automatic evaluation methods based on recall, precision and F1 scores. The results were encouraging, and we hope to do more investigations of automatic indexing in the future. More discussion of the results can be found in El Haj et al., 2012.

CESSDA-ELSST project

The aims of the CESSDA-ELSST project include: to revise and update both HASSET and ELSST; to convert both thesauri to a concept-based model; to review the relationship between the two thesauri; to convert ELSST to SKOS; to redevelop the thesaurus management applications; to streamline the management processes; and to review licence procedures.

---

13 http://www.gla.ac.uk/schools/critical/research/fundedresearchprojects/samuels/
14 http://brightstardb.com/
15 http://www.w3.org/2001/sw/wiki/Pubby

Revising and updating the thesauri

The project encompasses a thorough revision and update of both thesauri. This includes removing redundancy where it has occurred and making sure preferred terms have up-to-date labels. It also involves providing scope notes as often as possible, which is helpful not just to users, but also to the distributed team of translators who turn the Source Language terms in English into hierarchies in other languages. A review of the top terms of both thesauri is also planned – are there currently 298 in HASSET, 218 in ELSST, with a view to reducing their number and thus making the thesaurus easier to browse.

Converting thesauri to a concept-based model

In CESSDA-ELSST, the new international guidelines on thesaurus construction and use, ISO 25964 [16], are being followed as far as possible. ISO 25964 supports interoperability in two ways. Part 1 (ISO 24964-1) presents data in a standard way to enable import and use in other systems, while Part 2 (ISO 24964-2) covers the complementary use of vocabularies, including defining mappings between terms/concepts of one thesaurus and those of another (ISO 25964-2 press release, 2013).

The explicit data model presented in ISO 24964-1 clearly distinguishes between concepts and the terms used to represent the concepts, a distinction Dextre Clarke and Zeng (2012) argue is necessary for computer software to perform on the Semantic Web. Thus, a major part of the CESSDA-ELSST project was to move both HASSET and ELSST from a term-based to a concept-based model. Preferred terms become labels for concepts, and in a multilingual thesaurus like ELSST, different language versions of preferred terms are just alternative labels for the same concept.

Reviewing the relationship between the two thesauri

ISO 24964-2 also enabled us to define formally the relationship between HASSET and ELSST. Given that the two thesauri are so closely related, we began with the hypothesis that both thesauri could be merged.

The thesauri had grown apart over the years, so we reviewed all the terms and relationships that were in ELSST, not HASSET, and resolved the differences between the two thesauri wherever possible. Results of this alignment exercise suggested that, instead of forcing the two thesauri to merge, their common set of core concepts should be kept identical wherever possible, but allowed to diverge in clearly defined ways. Rather than being seen in terms of merging, their relationship can best be described in terms of a mapping. In this way, both thesauri can retain their integrity and identity.

The relationship between the two thesauri is captured through the use of axioms and constraints. Thus concepts that appear in both HASSET and ELSST ('core' concepts) must share the same Preferred Term label and BTs, but may vary in every other way (including, to a limited extent, in their scope notes, which allows for the inclusion of UK-specific information in HASSET but not ELSST).

---

16   http://www.niso.org/schemas/iso25964/

Converting ELSST to SKOS

The SKOS version of ELSST is based on the same technology as that of SKOS HASSET. Note that, like ISO 24964-1, SKOS is concept-based, and the development teams responsible for SKOS and ISO 25964 (respectively) have maintained a close working relationship throughout, leading to good compatibility between the standards (24964-2 Press release, 2013).

New thesaurus management system

Previously, the two thesauri were mounted on separate databases, which made them difficult to maintain: identical information had to be entered in two separate places, which led to errors and duplication of effort. The new thesaurus management system brings the two thesauri together within the same database structure. This is essential to managing them in an efficient manner. However, the two thesauri maintain separate user interfaces, supporting their different identities, audiences and, where appropriate, content. HASSET terms are linked to the studies at the UK Data Service that have been indexed with them, while ELSST provides a link to its multilingual equivalents. The user interfaces otherwise have identical features.

A novel feature of both the user and management interfaces is the implementation of a visualisation tool (see Figure 1) for navigation, in addition to the traditional tree structures.
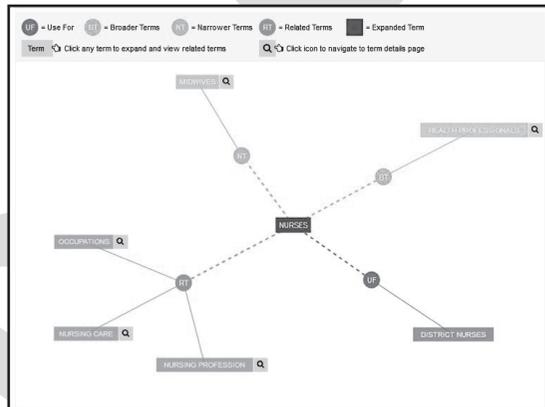

Figure 1

Visual graph view of the term NURSES:

Management permissions are controlled via Shibboleth, and authorised users can, as appropriate, suggest, discuss, and implement changes or translations, in the relevant thesaurus. Core concepts can be 'demoted' to non-core concepts in either ELSST or HASSET, and conversely, non-core concepts in HASSET can be 'promoted' to core concepts.

The thesaurus management system is now complete and ready to be used by both UK Data Service staff and by the international translators committee for ELSST, who are responsible for the translations of ELSST.

Review of licence procedures

The CESSDA-ELSST project has resulted in a new licence procedure. Both HASSET and ELSST are available free of charge to outside organisations for non-commercial purposes on completion of a licence. More work is planned to bring the licensing process online.

Conclusion and future work

The changing data landscape, especially the Semantic Web and Big Data, has brought new challenges for data archiving, in particular for the discovery, access, interoperability, integration, and annotation of often large and heterogeneous data. Thesauri have an important role to play in these areas. The two projects, SKOS-HASSET and CESSDA ELSST, have made good progress in positioning both HASSET and ELSST to meet these challenges. Both thesauri have been moved to a concept-based model, converted to SKOS, and the relationship between them has been clearly defined. The new thesaurus management system is expected to bring efficiencies for both users and developers. Experiments have also been undertaken on exploiting the thesauri for automatic indexing. In the next phase of the CESSDA-ELSST project, work will continue on the updating and reviewing the thesauri. Feedback on any aspect of the work is welcome either via the HASSET and ELSST web sites, or via the blog on the CESSDA-ELSST project page[17].

Acknowledgements

References

Bell, Darren (2012) *From Tuples to Triples: applying SKOS to HASSET – a technical overview*, SKOS-HASSET blog. Available at http://hassetukda.wordpress.com/2012/12/20/from-tuples-to-triples-applying-skos-to-hasset-a-technical-overview/

Berners-Lee, T., Hendler, J., and Lassila, O. (2001) 'The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities', *Scientific American*, 284(5), p. 34–43. Available at http://www.scientificamerican.com/article.cfm?id=the-semantic-web and http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html

Dextre Clarke, Stella, and Lei Zeng, Marcia (2012) 'From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling', *Information Standards Quarterly*, 24(1). Available at http://www.niso.org/publications/isq/2012/v24no1/clarke/

Dunsire, Gordon; Harper, Corey; Hillmann, Diane, and Phipps, Jon (2012),' 'Linked data vocabulary management: Infrastructure support, data integration, and interoperability', *Information Standards Quarterly*, 24(2/3). Available at http://www.niso.org/apps/group_public/download.php/9411/FE_Dunsire-etal_VocabMgmt_isqv24no2-3.pdf

El-Haj, Mahmoud; Balkan, Lorna; Barbalet, Suzanne; Bell, Lucy and Shepherdson, John (2013) 'An Experiment in Automatic Indexing Using the HASSET Thesaurus' *The 5th Computer Science and Electronic Engineering Conference (CEEC'13),* IEEE Xplore. 17-18 September 2013, University of Essex. doi: 10.1109/CEEC.2013.6659437 Available at http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6659437

ISO 25964-2 Press Release (2013): http://nkos.slis.kent.edu/PressReleaseISO25964-2.pdf

---

17   *https://elsst.wordpress.com/*

Méndez, Eva, and Greenberg, Jane (2012) 'Linked data for open vocabularies and Hive's global framework', *El Profesional de la Informacion*, May/June 2012, 21(3) p. 236. Available at http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/03_eng.pdf

Voß, Jakob (2007) 'Tagging, Folksonomy & Co – Renaissance of Manual Indexing?' In *Proceedings of the 10th International Symposium for Information Science*, Cologne, Germany. Available at http://arxiv.org/pdf/cs/0701072v2.pdf

W3C: SKOS/Datasets (2015) http://www.w3.org/2001/sw/wiki/SKOS/Datasets

WC3: Vocabularies (2015) http://www.w3.org/standards/semanticweb/ontology

ADVERTISING: Mail your publicity with the next issue. For £80 you can reach 350 library and information workers in government departments, universities, social services departments, voluntary associations, research organisations, professional bodies etc.

BACK ISSUES still available (£14 each):

Cheques payable to ALISS (payment with order)

Editor: Heather Dawson, British Library of Political and Economic Science, 10 Portugal Street, London WC2A 2HD. Email: h.dawson@lse.ac.uk