

# Catalogue and Index

Periodical of the Cataloguing and Indexing Group,  
a Special Interest Group of CILIP, the Library and Information  
Association



## Editorial

September 2018, Issue 192

Welcome to issue 192 of Catalogue and Index which is devoted to the topic of indexing. The term “indexing” can have different meanings, as indexing takes place in many different ways and formats. So, we are delighted that this issue gives a taste of the different ways indexing is used in various parts of the information professions. We have discussions about Library of Congress Subject Headings (LCSH) and FAST, which may be the ways many of us encounter indexing in everyday cataloguing life. However, the issue also contains articles about keywords in abstracting and indexing databases, subject indexing in repositories, and back-of-the-book indexing. We hear from those who do indexing and lead developments in indexing, as well as from those who research indexing. Above all, this issue shows the variety and vibrancy of indexes, indexing and indexers.

The issue starts with an article by Alan Danskin about subject indexing using FAST, including an analysis of how FAST is used at the British Library. Robert Kasenchak explores the topic of information discovery and retrieval through subject based classification and specialized language in repositories. Clare Playforth talks about using Library of Congress Classification to index theses in an institutional repository. Helen Bilton writes about indexing books, including the historical development of indexing and a discussion about how indexing is carried out in the present day. Paula Clarke Bain explores the use of humour in book indexes, analysing the jokes included in indexes for books about comedy. Rosemary Dear writes about the history and current-day activities of the Society of Indexers. Pilar Wyman describes the work and life of a book indexer, including the impact of technological changes to the indexing profession.



## Contents

3-8: FAST forward? By Alan Danskin

9-16: Methodologies for classifying repositories, by Robert T. Kasenchak

17-19: Subject indexing in a repository, by Clare Playforth

20-23: History of index creation, by Helen Bilton

24-30: Comedic indexes, by Paula Clarke Bain

31-32: Foundation of the Society of Indexers, by Rosemary Dear

33-36: Directions in indexing, by Pilar Wyman

37-39: Changed needed in the subject index, by Liz Antell

40-47: Online language classification, by Lucile Deslignères

48-51: Machine translation, by Lynne Bowker

52-57: Retrospective cataloguing Case study, by Max Zanotti

58-60: CILIP Conference 2018 report, by Lynn Thorn

61: Letters & responses

Liz Antell discusses ethical issues in LCSH, using a specific example of a race-related term to analyse the harm that is caused by not dealing with historical, unethical headings that still lurk in our catalogues. Lucile Deslignères outlines the creation of an index to organise resources on the website of the Language Centre at the University of Oxford, illustrating the challenges of categorising languages. Lynne Bowker analyses the issues of language and translation of keywords, in the context of indexing resources in abstracting and indexing databases.

The final section includes a case study, conference report and a letter, all focussing generally on cataloguing and classification. Max Zanotti provides a report about a retrospective cataloguing project at Chatham House Library, which was compiled for his master's degree. Next, Lynn Thorn, winner of a CIG bursary to attend the CILIP conference in July, reports on this conference, including reference to cataloguing and indexing related themes. Finally, Natasha Aburrow-Jones writes a short response to an article in issue 191 of Catalogue and Index. We really welcome such responses, so if you have anything to say about any of this issue's articles, please write to the editors.

Karen F. Pierce: [PierceKF@Cardiff.ac.uk](mailto:PierceKF@Cardiff.ac.uk)

Deborah Lee: [Deborah.Lee@courtauld.ac.uk](mailto:Deborah.Lee@courtauld.ac.uk)

FAST stands for Faceted Application of Subject Terms.<sup>1</sup> It is an OCLC research project development of which began in 1998, in collaboration with Library of Congress. The objective was to combine the rich vocabulary of LCSH with a simplified syntax to deliver a schema that is easier for cataloguers and indexers to understand, control and apply but also enables effective subject access. The schema maintains upward compatibility with LCSH enabling any valid LCSH subject headings to be converted to FAST. FAST headings have been automatically generated in OCLC records and are already a familiar presence in catalogue records.

FAST is structured around 8 facets:

Topical; Geographic; Chronological; Events; Names as Subject; Name-titles; Form-genre

As well as LCSH terms, FAST includes headings from NACO and LCGFT to facilitate cataloguing efficiency. There are still some elements of pre-coordination in FAST, but subdivisions may only be assigned within a FACET, as refinements of a terms; subdivisions may not cross facets, as in LCSH.

FAST headings are widely distributed in MARC 21 records and can be accommodated in other schema, including Dublin Core. FAST is also available as linked data under an Open Data Commons Attribution License (ODC-BY). Application is supported by the “searchFAST” and “assignFAST” services.

FAST has been adopted by a number of research libraries in the US and by other services. Users who contributed to an OCLC report in 2013, included National Library of New Zealand, Bodleian Libraries, University of Oxford, University of Amsterdam. The scope of application ranged from general subject indexing of the collection to specific projects or classes of material.

In 2016 the British Library consulted with our stakeholders about our future choice of subject standards.<sup>2</sup> The survey posed three scenarios for comment:

1. The British Library proposes to adopt FAST selectively to extend the scope of subject indexing of current and legacy content.
2. The British Library proposes to implement FAST as a replacement for LCSH in all current cataloguing, subject to mitigation of the risks identified in the background paper; in particular, the question of sustainability.
3. The British Library proposes to implement Abridged DDC selectively to extend the scope of subject indexing of current and legacy content.

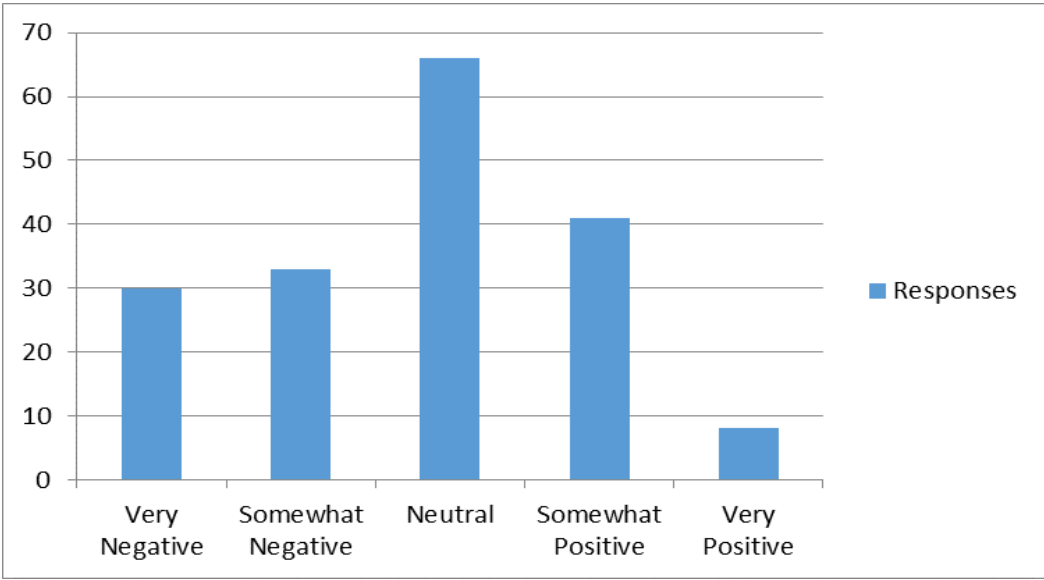
A total of 60 responses was received. Respondents were not required to specify their location, those who did were mainly from UK (21/33) with others (6/33) from USA, Ireland (3/33) and sole respondents from three other countries.

---

1. OCLC Research. FAST (Faceted Application of Subject Terminology) <https://www.oclc.org/research/themes/data-science/fast.html>

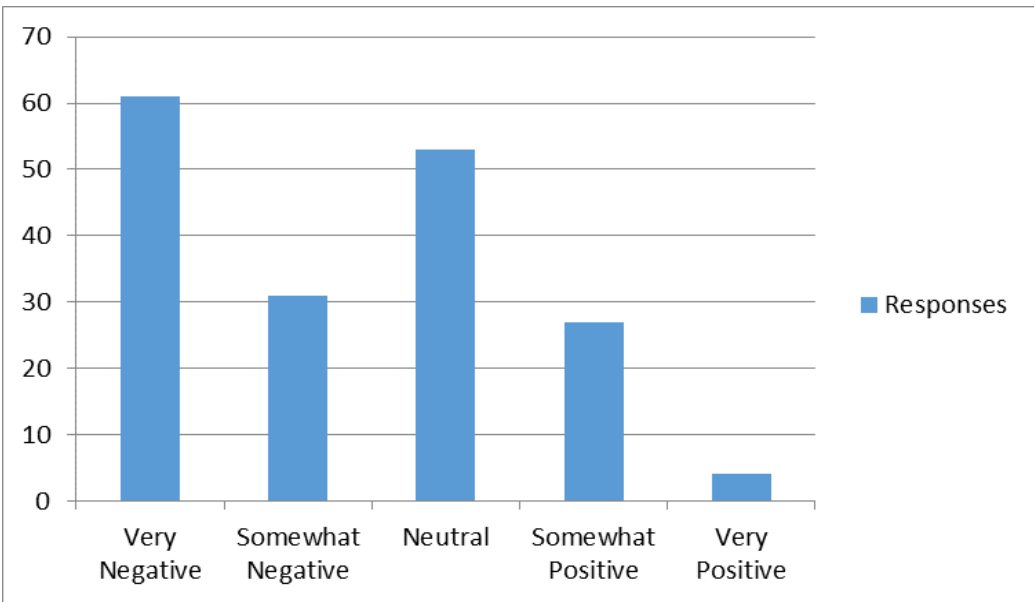
2. British Library. Consultation on Subject Indexing and Classification standards applied by the British Library. February 2016. <http://www.bl.uk/bibliographic/pdfs/british-library-consultation-fast-abridged-dewey.pdf>

Respondents were required to rate their reaction to each proposal from 1 (very negative) to 5 (very positive). The range of responses to the two questions about FAST is interesting.



Proposal 1 The British Library proposes to adopt FAST selectively to extend the scope of subject indexing of current and legacy content

The responses to the proposal to use FAST selectively indicated significant hostility to the idea. The idea also had its supporters, although they were less vehement than the opponents. The largest group was neutral.



Proposal 2 The British Library proposes to adopt FAST selectively to extend the scope of subject indexing of current and legacy content

The proposal to replace LCSH with FAST was much more polarising because it would have a much greater impact on stakeholders who use BNB records, including CIP, for copy cataloguing or who use our free Z39.50 service to derive MARC records. There was considerable hostility to FAST and several respondents were sceptical about its utility for supporting subject search. More nuanced responses acknowledged that FAST could be more efficient to apply but were concerned about its sustainability. FAST was an OCLC project. There was little doubt expressed that LC is committed to LCSH but what guarantee was there of OCLC's commitment to FAST?

In its response, the Library acknowledged the importance of sustainability and undertook to continue its discussions with OCLC with regard to sustainability and to further evaluate the efficiency and quality of FAST.<sup>3</sup>

Before describing the outcome of this work, I want to explain the context in which the British Library is considering replacement of LCSH. This is not a decision that we would take lightly. LCSH has been our main subject indexing scheme since 1994, when it was reinstated to replace COMPASS (Computer Aided Subject System). COMPASS was introduced as a response to the rising cost of applying both PRECIS and LCSH to the British National Bibliography but itself proved unsustainable. However, having a “*main subject indexing scheme*” is not the same as having one subject indexing scheme.

One of the objectives of our metadata strategy is to improve overall efficiency by reducing the number of metadata standards in use. The Library's Everything Available portfolio has an objective to improve discovery generally and among its other aspirations is an improved capability to visualise the collection. At present, the patchy coverage of subject indexing and the variety of standards in use make a subject view of the collection very patch in deed. The following graph benchmarks current support for subject indexing against legacy records and against catalogue records in our Aleph ILS.

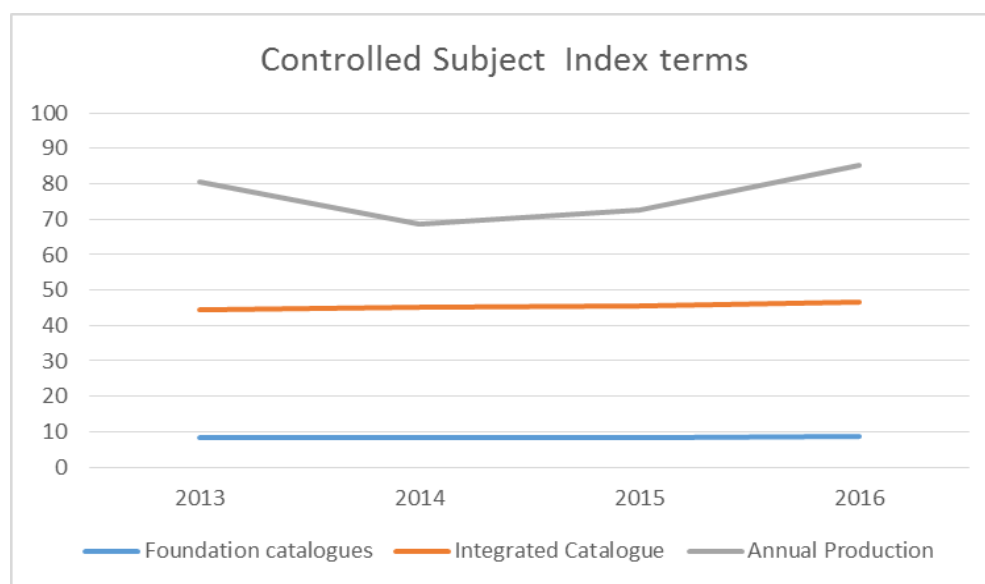


Figure 1 Controlled Subject Index Terms in catalogue production db

3. British Library. Response to the consultation on subject standards. 18/7/2016 <http://www.bl.uk/bibliographic/pdfs/british-library-response-survey-subject-standards.pdf>

The annual audit of catalogue data in the Aleph ILS production data allows us to compare metrics for a number of metadata attributes across three data sets: Annual production (data created in the production year preceding the audit); Foundation catalogues (retrospectively converted catalogue records from our foundation collections, e.g. British Museum Catalogue of printed books); Integrated catalogue (all records in the Aleph production database which are exported to Explore the British Library).

You can see that the overall figure represented by the red line hovers around the 45% mark. So fewer than half of our catalogue records in Aleph have controlled subject index terms. To put it another way, over 8 million records need subject index terms. About half of these are the old British Museum records, the remainder are primarily document supply records. Generally over 70 % of current records receive controlled subject terms – this dipped recently as a result of the influx of e-books under non-print legal deposit, but is improving following the implementation of the batch upgrade programme. There are a number of initiatives underway to address problem of subject indexing, but there is a fundamental problem of capacity.

LCSH is not sufficiently scalable to: 1) achieve 100% of coverage of the current intake of trade literature; 2) to address deficiencies of legacy data; 3) to extend subject indexing to workflows in other collections that do not apply a scheme.

The Library has run a number of projects to test application of FAST. These have involved both experienced LCSH indexers and staff with no prior experience of LCSH. FAST has been applied to a wide range of materials, including academic monographs in a variety of European languages; official publications and grey literature; e-journals; databases; uncatalogued Asian and African collection material. Feedback from staff participating in the projects has been positive. FAST was found to be intuitive and easy to use and quickly attained confidence in its application. The tools, such as searchFAST, were viewed very positively. The overall outcome was very positive, as these were projects for which we could not have resourced the application of LCSH.

The complexity of LCSH's application is the main barrier to extending its use in the Library. Competence in application is achieved only after intensive training and a relatively lengthy period of supervision. Short term projects, which have limited funding, recruit staff on fixed term contracts ranging from a few months to at most two years. There is intense pressure for these project staff or interns to become productive as quickly as possible. Staff are recruited for their specialist language or subject knowledge and may have little cataloguing experience. They generally use Excel input forms to record attributes which are converted to MARC for load to Aleph. Under these circumstances it is unrealistic to devote weeks of training to LCSH, but they can be relatively quickly trained in the essentials of subject analysis and the application of FAST.

For staff who are experienced in the application of LCSH we found little difference in the speed of subject indexing, but it is difficult to increase this pool of experience staff, many of whom are nearing retirement. Is LCSH sustainable without a pool of expert trainers? UK publishing, legal deposit intake and mass digitization are increasing the number of things that require indexing. In response our processes are becoming increasingly automated and reliant on external sources of metadata. The complex syntax of LCSH is not well suited to machine processing and it seems likely that a faceted system such as FAST would be more adaptable to these challenges.

Our analysis also considered the qualitative aspects of FAST. It must be kept in mind that from the perspective of catalogue users breadth of indexing may be as, or more important than depth of indexing. Nevertheless, our analysis has not revealed any inherent qualitative difference. FAST can be applied at the same level of specificity as LCSH, but the level can be varied by local application rules. This makes FAST more flexible than LCSH, but can cloud perceptions of relative quality.

There are therefore many compelling reasons why FAST could be considered a viable successor to LCSH, but it would represent a significant risk to proceed without confidence in the sustainability of FAST. The Library and other institutions have been working with OCLC and others to discuss what a FAST service would look like and how it would be governed in the interests of the community of users.

In April 2016, we discussed our interest in “production-izing” FAST with Mary Sauer-Games, Vice President Project Management and Product Marketing. Contact with OCLC continued during 2016-17 and we also engaged informally with the so-called “FAST five” academic libraries in the US: Cornell, Harvard, Yale, Brown, and Columbia. This group initiated a survey in late 2017, whose findings were published in February 2018.<sup>4</sup> The survey posed 14 questions and garnered 586 responses from institutions predominantly in US, UK, Canada, Australia and New Zealand (95%). Most respondents (85%) were working with metadata and the majority were practitioners (57%); 21% identified as administrators or supervisors. Headline findings were that 30% of respondents were already using FAST and a further 7% planned to do so in future. Over half (57%) of those not using FAST stated a preference for pre-coordinated subject headings. The balance was evenly split between those who could not use FAST because of system constraints and those deterred by its project status. The most commonly cited benefit (57%) respondents perceived, is that FAST enables provision of subject access to records that would otherwise lack it; improved discovery in faceted environment (55%) and efficiency (51%) also scored highly; many other benefits were cited including reduced training time; expanded pool of subject indexers, potential for crowdsourcing and support for linked data applications.

Respondents were also asked to rank the features and enhancements they would like to see in a FAST service:

- Mechanism to support ongoing maintenance: 121
- Production tool for FAST heading look-up and seamless addition to record at point of cataloguing: 108
- Ability to request new FAST headings: 89
- Ability to generate FAST from LCSH strings automatically when doing original cataloguing or enhancing a record in Connexion: 87
- Current SearchFAST (the search interface that simplifies heading selection): 86
- Current algorithmic addition of FAST headings with the identifier in the \$0 to WorldCat records with LCSH: 81
- Ability to create and submit new FAST headings: 80
- Current FAST converter (a web application that converts LCSH to FAST): 74
- Batch conversion of local LCSH to FAST: 70
- Current functionality in OCLC WorldShare Record Manager to add FAST with the “text” view in editor: 38
- More robust FAST headings for medical areas: 30

It was concluded from this that FAST has a substantial user base who would be interested in a production version and that establishing a service would have the effect of increasing adoption and usage.

The British Library also presented our own research at conferences in US and Germany and published an article in *Cataloguing & Classification Quarterly*.<sup>5</sup>

---

4. Survey on “Next steps for FAST”: summary of results, conclusions and next steps. February 8, 2018. <https://docs.google.com/document/d/1HdPrWOKGprYKRSVO9JYiaQ6DULpdDjek3P3C1V7nXqs/edit> (Report)  
Kate Harcourt. “Next steps for FAST” ALA Midwinter Conference February 12, 2018 (Slides)  
<https://docs.google.com/presentation/d/1XMiaj91qsTKFQlrhRaWHpxyHAPItR-ixIFpR7dbZtLw/edit#slide=id.p3>

5. Janet Ashton & Caroline Kent (2017) New Approaches to Subject Indexing at the British Library, *Cataloguing & Classification Quarterly*, 55:7-8, 549-559, DOI: 10.1080/01639374.2017.1354345 <https://doi.org/10.1080/01639374.2017.1354345>



In late 2017, OCLC initiated a series of online meetings with interested parties to understand community workflows and needs. In response to this invitation the Library specified its own requirements, including commitment to a “*product*” not a “*project*” and outlined some ideas for governance and future development. As a result of these various inputs, OCLC is actively working to transition the FAST service from an experimental research environment to production servers. This was initially expected to be completed by September 2018, but is now scheduled to be completed by March 2019. OCLC’s Jody DeRidder (Director Metadata Frameworks) gave an update to the Faceted Subject Access Interest Group meeting at ALA Annual, in New Orleans on 23rd June 2018, and asked for feedback on governance, scope and community relationship.

OCLC has partnered with representatives from the library community to create a new FAST Policy and Outreach Committee (FPOC). FPOC is an advisory and outreach body that will represent users of FAST for the purposes of:

- establishing editorial policies regarding terms
- overseeing community engagement, term contributions and procedures
- recommending directions and goals for development/improvements
- and much more (see Regulations for the Conduct of Committee Business; attached)

Membership of the committee will be a mix of OCLC liaisons and appointive members. Appointive members for the transitional first term will be nominated by institutions already committed to FAST development. During this transitional phase terms of membership will be staggered, with members serving for either 1 year or 2 years. Subsequently, the standard term will be two years and responsibility for appointing members will devolve to the committee and OCLC. Nominations will then be open to all community members and other interested communities will also be able to make nominations.

Among the other initiatives being explored is the deployment of a sandbox to facilitate community participation in vocabulary development. The British Library will also meet with OCLC and other stakeholders at Library of Congress in October, to discuss the implications of FAST development for the current relationship with LCSH. The FAST community wants more flexibility to create new terms, but a significant strength of FAST is the common vocabulary it shares with LCSH. There are also questions of scope. To what extent should FAST replicate facets that Library of Congress has extracted from LCSH and published as linked data, e.g. LCGFT?

In conclusion, there is progress on developing FAST as a service. The British Library will continue to extend the use of FAST to provide subject access to content that would not otherwise be indexed. Following the discussions at Library of Congress in October, we will review the future relationship between FAST and LCSH. No decisions about the future of LCSH and BNB will be taken without consultation with stakeholders and the UK cataloguing community.



### Abstract

Large, specialized repositories of content present unique problems for information discovery and retrieval: the sheer size of the repositories, the specialization (and corresponding demanding expectations) of the researcher-users, and the specialized language—including many acronyms and other abbreviations—found in any sufficiently advanced field of research. No two collections cover identical subject matter; what is the best practice for classifying such collections to optimize retrieval?

### What is a “Special Collection”?

By *special collections or repositories* I mean digital collections held by libraries, publishers, aggregators, and other organizations.

Some such collections cover a broad topic, for example, *PLOS One* or *Science Magazine*, both of which publish broadly on the topic of science, albeit with different emphases. Other collections have more specific topical foci, such as those held by society and association publishers; topics can be as broad as physics or chemistry, or as focused as optics or cancer research. Still others are a subset of a larger holding of content; many large university libraries have region-based collections, papers of local politicians, theses and dissertations, and other specialized collections.

The size of these collections ranges from the tens of thousands to the millions of content items. Below are some representative examples:

JSTOR: over 12,000,000 items, including about 7 million research articles

IEEE: over 4,000,000 items (research articles, proceedings, and other content)

PLOS One: over 300,000 research articles

American Institute of Physics: over 900,000 items

University of Florida Electronic Theses and Dissertations: approximately 30,000 items

### Special Collections and Discovery

The basic tenets of information science have long prescribed subject-based classification—indexing the content using a knowledge organization system (KOS) such as a taxonomy, thesaurus, ontology, or other controlled vocabulary—as a pillar of aiding content discovery and retrieval; this is even more important for large digital collections.

Simple, free-text searches of large repositories are simply ineffective; the ambiguities of language and variations in usage are prohibitive barriers to effective retrieval in large collections. This is compounded by the twin problems of ambiguous words (homophony and polysemy) and synonymy.

## Failings of Free-Text Searching

The goal of good information retrieval is to deliver the user *all* of the relevant information on a topic as well as *only* the relevant information. No researcher wants to look through tens of thousands of results for a search on the text string “mercury” to find the content about astronomy (and not about cars, chemical elements, or Roman gods).

The screenshot shows a Google Scholar search interface. At the top, the Google Scholar logo is on the left, and a search bar contains the word "mercury" with a magnifying glass icon on the right. Below the search bar, it says "Articles" and "About 2,890,000 results (0.03 sec)". On the left side, there are filters: "Any time" (with sub-options: Since 2018, Since 2017, Since 2014, Custom range...), "Sort by relevance", "Sort by date", checkboxes for "include patents" and "include citations", and a "Create alert" button. The main results area shows several entries:

- [CITATION] 2.0-new features for the visualization and investigation of crystal structures Macrae  
CSD Mercury - Clare F  
☆ 99 Cited by 532 Related articles
- Mercury** CSD 2.0—new features for the visualization and investigation of crystal structures  
CF Macrae, JJ Bruno, JA Chisholm... - Journal of Applied ..., 2008 - scripts.iucr.org  
The program **Mercury**, developed by the Cambridge Crystallographic Data Centre, is designed primarily as a crystal structure visualization tool. A new module of functionality has been produced, called the Materials Module, which allows highly customizable searching of ...  
☆ 99 Cited by 5514 Related articles All 4 versions
- Mercury**: visualization and analysis of crystal structures  
CF Macrae, PR Edgington, P McCabe... - Journal of Applied ..., 2006 - scripts.iucr.org  
Since its original release, the popular crystal structure visualization program **Mercury** has undergone continuous further development. Comparisons between crystal structures are facilitated by the ability to display multiple structures simultaneously and to overlay them ...  
☆ 99 Cited by 4716 Related articles All 5 versions
- [PDF] Stratigraphy and geologic history of **Mercury**  
PD Spudis, JE Guest - **Mercury**, 1988 - academia.edu  
The history of **Mercury** can be reconstructed through photogeological analysis of global geologic relations of rock-stratigraphic units. Mercurian stratigraphy is subdivided into five time-stratigraphic systems based on mappable basin and crater deposits. These systems ...  
☆ 99 Cited by 159 Related articles All 7 versions
- Determination of submicrogram quantities of **mercury** by atomic absorption spectrophotometry  
WR Hatch, WL Ott - Analytical Chemistry, 1968 - ACS Publications  
The procedure outlined describes an extremely sensitive and accurate method for the determination of **mercury** down to 1.0 ppb in solution. This procedure has been applied to nickel and cobalt metal as well as rock samples and soil samples containing organic ...  
☆ 99 Cited by 1583 Related articles All 3 versions
- Atmospheric **mercury**—an overview  
WH Schroeder, J Munthe - Atmospheric environment, 1998 - Elsevier  
This paper presents a broad overview and synthesis of current knowledge and understanding pertaining to all major aspects of **mercury** in the atmosphere. The significant physical, chemical, and toxicological properties of this element and its environmentally ...  
☆ 99 Cited by 1707 Related articles All 8 versions

Note that the *Google Scholar* search shown above contains some 2.8 million results, of which the first six shown contain articles about both planetary science and elemental chemistry.

Conversely, forcing users to brainstorm every possible variant and synonym of a concept is not an avenue conducive to effective research (or delighted users, or repeat customers, or renewed subscribers).

Any time

Since 2018

Since 2017

Since 2014

Custom range...

Sort by relevance

Sort by date

☒ include patents

☒ include citations

☒ Create alert

Evaluation of diagnostic criteria for ankylosing spondylitis

..., HA Valkenburg, A Cats - Arthritis & ..., 1984 - Wiley Online Library

Abstract The New York and the Rome diagnostic criteria for ankylosing spondylitis (AS) and the clinical history screening test for AS were evaluated in relatives of AS patients and in population control subjects. The New York criterion of pain in the (dorso) lumbar spine lacks ...

☆ 99 Cited by 5165 Related articles All 16 versions

European guidelines on cardiovascular disease prevention in clinical practice: third joint task force of European and other societies on cardiovascular disease ...

..., I Graham, G Mancia, VM Cats... - European heart ..., 2003 - academic.oup.com

... Ole Faergeman. Search for other works by this author on: Oxford Academic. PubMed. Google Scholar. Ole Faergeman Ian Graham. Search for other works by this author on: Oxford Academic. PubMed. Google Scholar. Ian Graham Giuseppe Mancia. Search for other works by this author ...

☆ 99 Cited by 3008 Related articles All 19 versions

The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy

..., B Huitfeldt, B Amor, A Calin, A Cats... - ... : Official Journal of ..., 1991 - Wiley Online Library

Classification criteria for most of the disorders belonging to the spondylarthropathy group already exist. However, the spectrum of spondylarthropathy is wider than the sum of these disorders suggests. Sero-negative oligoarthritis, dactylitis or polyarthritis of the lower ...

☆ 99 Cited by 2570 Related articles All 16 versions

Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer

J Tol, M Koopman, A Cats, CJ Rodenburg... - ... England Journal of ..., 2009 - Mass Medical Soc

Background Fluoropyrimidine-based chemotherapy plus the anti-vascular endothelial growth factor (VEGF) antibody bevacizumab is standard first-line treatment for metastatic colorectal cancer. We studied the effect of adding the anti-epidermal growth factor receptor ...

☆ 99 Cited by 1350 Related articles All 14 versions

An overview of the incidences and costs of low back pain.

JW Frymoyer, WL Cats-Baril - The orthopedic clinics of North ..., 1991 - europepmc.org

The basic premise of this article is that low back disorders are extremely prevalent in all societies, and probably have not increased substantially over the past two decades. What has increased is the rate of disability, the reasons for which are uncertain. Not only has this ...

☆ 99 Cited by 1208 Related articles All 7 versions

Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations.

JL Van Saase, LK Van Romunde, A Cats... - Annals of the ..., 1989 - ncbi.nlm.nih.gov

The prevalence of mild and severe radiological osteoarthritis was investigated in a random



Google Scholar

felines

Articles

About 32,200 results (0.07 sec)

Any time

Since 2018

Since 2017

Since 2014

Custom range...

Sort by relevance

Sort by date

☒ include patents

☒ include citations

☐ Create alert

Oral infections with Toxoplasma cysts and oocysts in **felines**, other mammals, and in birds

NL Miller, JK Frenkel, JP Dubey. - The Journal of parasitology, 1972 - JSTOR

Toxoplasma oocysts were shed by 1 out of 7 bobcats (*Lynx rufus*), 1 mountain lion (*Felis concolor*), and 2 Asian leopard cats (*Felis bengalensis*) after the ingestion of toxoplasma cysts. Oocysts were not shed by 16 nonfeline mammals from 7 orders nor by 9 birds from 5 ...

☆ Cited by 299 Related articles All 5 versions

Structural studies of the phase separation in the UO<sub>2</sub>-PuO<sub>2</sub>-Pu<sub>2</sub>O<sub>3</sub> ternary system

T Truphémus, RC Belin, JC Richaud... - Journal of Nuclear ..., 2013 - Elsevier

In the oxygen hypo-stoichiometric range of (U 1- y Pu y) O 2- x mixed oxide MOX fuels, the U-Pu-O phase diagram is known to exhibit a large biphasic domain depending on the Pu content. However, the phase equilibria are still to be fully described as various ...

☆ Cited by 32 Related articles All 5 versions

[CITATION] Jaguar predation and conservation: cattle mortality caused by **felines** on three ranches in the Venezuelan Llanos

R Hoogesteijn, A Hoogesteijn, E Mondolfi - Symposium of the Zoological Society of ..., 1993

☆ Cited by 136 Related articles

[HTML] Who ate whom? Adaptive Helicobacter genomic changes that accompanied a host jump from early humans to large **felines**

M Eppinger, C Baar, B Linz, G Raddatz, C Lanz... - PLoS ..., 2006 - journals.plos.org

Helicobacter pylori infection of humans is so old that its population genetic structure reflects that of ancient human migrations. A closely related species, Helicobacter acinonychis, is specific for large **felines**, including cheetahs, lions, and tigers, whereas hosts more closely ...

☆ Cited by 165 Related articles All 15 versions

Quantity discrimination in **felines**: a preliminary investigation of the domestic cat (*Felis silvestris catus*)

PE Pisa, C Agrillo - Journal of Ethology, 2009 - Springer

A large body of studies has investigated the capacity of non-human primates, dogs, birds and lower vertebrates to estimate different quantities of objects or events. Little attention, however, has been devoted to **felines**, and no study has specifically concentrated on cats' ...

☆ Cited by 106 Related articles All 11 versions

Beyond Mead: Symbolic interaction between humans and **felines**

JM Alger, SF Alger - Society & Animals, 1997 - booksandjournals.brillonline.com

Recent research on the cognitive abilities and emotional capacities of animals has fueled the animal rights movement and renewed debate over the differences between human and non-human animals. This debate has not been central to sociology, although George ...

The examples above illustrate the failings of free-text-only searching: this process considers neither synonymy nor polysemy; it simply searches for words in text. In some cases, not even regular language variants (such as simple English plural forms) are accounted for:

Google Scholar

horse

Articles

About 3,200,000 results (0.03 sec)

Any time

Since 2018

Since 2017

Since 2014

Custom range...

Sort by relevance

Sort by date

☒ include patents

☒ include citations

☐ Create alert

[HTML] Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface

K Red-Horse, Y Zhou, O Genbacev... - The Journal of ..., 2004 - Am Soc Clin Investig

Trophoblasts, the specialized cells of the placenta, play a major role in implantation and formation of the maternal-fetal interface. Through an unusual differentiation process examined in this review, these fetal cells acquire properties of leukocytes and endothelial ...

☆ 99 Cited by 565 Related articles All 12 versions

The historical trauma response among natives and its relationship with substance abuse: A Lakota illustration

MYHB Heart - Journal of psychoactive drugs, 2003 - Taylor & Francis

Historical trauma (HT) is cumulative emotional and psychological wounding over the lifespan and across generations, emanating from massive group trauma experiences; the historical trauma response (HTR) is the constellation of features in reaction to this trauma ...

☆ 99 Cited by 615 Related articles All 7 versions

Coronary arteries form by developmental reprogramming of venous cells

K Red-Horse, H Ueno, IL Weissman, MA Krasnow - Nature, 2010 - nature.com

Coronary artery disease is the leading cause of death worldwide. Determining the coronary artery developmental program could aid understanding of the disease and lead to new treatments, but many aspects of the process, including their developmental origin, remain ...

☆ 99 Cited by 360 Related articles All 16 versions

International competition

YHPIN KOREA - 1993 - sarp.org.pl

" Lets Run Park Yeongcheon," a theme park about horses and racetrack, will be developed on a 1,474,883-m2 site and is intended to become a local attraction that will benefit the local economy. It will improve the overall image of horse racing and contribute to the horse ...

☆ 99 Cited by 214 Related articles All 3 versions

[BOOK] Final report

HRIT Panel, E Buchanan - 2012 - archive.nefmc.org

Discussion: Mr. Brogan expressed concern about what happens when there is an external operating constraint and there are very limited or no funds. Mr. Calomo felt that the funds for observer coverage could be generated through political means. Mr. Mullen expressed ...

☆ 99 Cited by 241 Related articles All 2 versions

Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach

R Ghanem, J Red-Horse - Physica D: Nonlinear Phenomena, 1999 - Elsevier

This paper presents an efficient procedure for characterizing the solution of evolution equations with stochastic coefficients. These typically model the behavior of physical

The screenshot shows the Google Scholar interface with the search term 'horses'. The results list several articles, including 'Horses' by GG Simpson, 'Veterinary Medicine E-Book: A textbook of the diseases of cattle, horses, sheep, pigs and goats' by OM Radostits et al., 'Gait and the energetics of locomotion in horses' by DF Hoyt and CR Taylor, 'Leukoencephalomalacia in two horses induced by oral dosing of fumonisin B1' by TS Kellerman et al., and 'Heterogeneous catalysts for liquid-phase oxidations: Philosophers' stones or Trojan horses?' by RA Sheldon et al. Each entry includes a star icon, a citation count, and links to related articles and all versions.

**Articles** About 1,730,000 results (0.03 sec)

**Any time**  
 Since 2018  
 Since 2017  
 Since 2014  
 Custom range...

**Sort by relevance**  
 Sort by date

☒ include patents  
☒ include citations

☒ Create alert

**Horses.**  
 GG Simpson - [Horses.](#), 1951 - cabdirect.org  
 The sub-title of Professor Simpson's book-The Story of the **Horse** Family in the Modern World and Through Sixty Million Years of History-speaks for itself. This lucid, nontechnical account of the origin and development of the Equidae equidae Subject Category: Organism ...  
 ☆ 99 Cited by 410 Related articles All 2 versions

**[book] Veterinary Medicine E-Book: A textbook of the diseases of cattle, horses, sheep, pigs and goats**  
 OM Radostits, CC Gay, KW Hinchcliff, PD Constable - 2006 - books.google.com  
 The new 10th edition of this well-respected reference includes thoroughly revised and reviewed information with content on new diseases for the most up-to-date information on diseases effecting sheep, pigs, goats, **horses**, and cattle. Divided into two sections, General ...  
 ☆ 99 Cited by 2305 Related articles All 4 versions

**Gait and the energetics of locomotion in horses**  
 DF Hoyt, CR Taylor - [Nature](#), 1981 - nature.com  
 It seems reasonable that quadrupeds should change gait from a walk to a trot to a gallop in such a way as to minimize their energy consumption, as human beings are known 1 to change from a walk to a run at a particular speed (2.4 ms<sup>-1</sup>) below which walking requires ...  
 ☆ 99 Cited by 1045 Related articles All 7 versions

**Leukoencephalomalacia in two horses induced by oral dosing of fumonisin B1.**  
 TS Kellerman, WF Marasas, PG Thiel... - [journal of veterinary ...](#), 1990 - europepmc.org  
 Leukoencephalomalacia (LEM) was induced by the oral administration of fumonisin B1 (FB1) to 2 **horses**: a filly received 59.5 mg/kg of a 50% preparation of FB1, administered in 21 doses of 1.25-4 mg/kg over 33 days; a colt, 44.3 mg/kg of 95% pure FB1 in 20 doses of 1 ...  
 ☆ 99 Cited by 596 Related articles All 8 versions

**Heterogeneous catalysts for liquid-phase oxidations: Philosophers' stones or Trojan horses?**  
 RA Sheldon, M Wallau, IWCE Arends... - [Accounts of Chemical ...](#), 1998 - ACS Publications  
 In the early seventies one of us<sup>1</sup> was involved in the development of the heterogeneous Ti (IV)/SiO<sub>2</sub> catalyst which forms the basis of the Shell process for the epoxidation of propylene with ethylbenzene hydroperoxide (reaction 1). 2 Halcon3 and ARCO4, 5 workers had ...  
 ☆ 99 Cited by 927 Related articles All 5 versions

**[book] Large animal internal medicine: diseases of horses, cattle, sheep, and goats**  
 BP Smith - 1996 - [jfrpcgfh01.storage.googleapis.com](#)  
 6Ckw0VDEt UP2JIPZkd 7QUQRDpm4 1yIAJ8KEM Fr1eFMwTA tAnnCm6eZ R0vOnMVjG IM3bkUU5Y BYL3vuMKe 3jMXUitFNk 4ZrNcrzRP gYBZhqCph gloR2HHmR xO6MjpiKi pzcYgUiil 4KTZhqZ51

## Classification Systems for Information Retrieval in Special Collections

It is well established (Jackson, 1971; Coates, 1988; Jain and Wadekar, 2011) that classification aids retrieval. Is it more effective to use some standard and readily available classification scheme (such as the Dewey Decimal System (DDC) or the Library of Congress Subject Headings (LCSH)) or a custom scheme designed for a specific collection? Or are author-supplied—or other uncontrolled—keywords just as effective?

### Author-Supplied Keywords

Generally speaking, author-supplied keywords are not useful for retrieval in large repositories. As authors of research articles (with the exception of library and information scientists) are not trained in the basic tenets of information tagging and retrieval, the keywords they supply do not follow good indexing practices (Janda, 2014).



Often, they are too broad to be useful for retrieval (“biology”) or heedless of the ambiguity of language (“evolution”). Regardless, without reference to a standardized vocabulary including synonyms and other variants, such inexpertly applied keywords are unhelpful for retrieval.

### Existing Vocabularies or Subject Headings

There are many vocabularies published under open-source (and other free-to-access) licenses (see for example [www.bartoc.org](http://www.bartoc.org) for a useful repository). One (or more) of these can be good starting points for custom vocabulary construction. However, existing vocabularies seldom match precisely the coverage of any given special collection; for example, using the Library of Congress Subject Headings (LCSH) to index any subject-specific collection will provide both far more subjects than required for the collection and not enough granularity in the subject-specific area to index the content thoroughly enough for specialized researchers and accurate information retrieval.

In addition, as special collections tend to be very large, it is increasingly common to require some kind of automatic indexing or classification for special collections; they are simply too large to index each content item by hand. Subject-heading style vocabularies (for example, again, LCSH) are particularly ill-suited to automatic classification (as they are expressed in the logic of subject headings instead of natural language phraseology) without substantial pre-processing rule and/or training sets. The pre-coordinated nature of subject-heading entries do not often match the language used in writing, making one-to-one matches of text to subject heading difficult.

For example, consider the LCSH Subject Heading:

*American drama--Afro-American authors*

It is difficult to imagine this phrase (complete with punctuation, and considering the archaic language in the subject heading) appearing as such in a modern research paper; therefore, mappings of natural language to subject heading concepts (or substantially curated training sets) is required for any kind of automatic categorization.

### Collection-Specific Vocabularies

The most useful practice for the indexing and cataloging of special collections for retrieval is to use collection-specific vocabularies (Hedden, 2010). As no two collections cover identical subject matter, the breadth and depth of a vocabulary required to index any given collection is unique. However, existing vocabularies can make excellent starting points for the construction of customized, collection-specific vocabularies. As noted above, many vocabularies are published under various open-source licenses; other vocabularies can be licensed or otherwise borrowed with permission.

To adapt an existing vocabulary to be suitable for indexing a special collection, it is necessary to both remove unnecessary terms/branches (describing subjects not included in the collection) and to augment the vocabulary to include subjects not included in (and, often, in more granular detail than described by) the starting vocabulary. This can be achieved by a number of means, including text and data mining (TDM) operations on the collection to be indexed and reviewed by subject matter experts. Terms derived from search logs and other forms of tracking user behavior can also be useful additions.

Using collection-specific controlled vocabularies to index and surface content in special collections can also provide useful features for discovery interfaces. Notably, providing type-ahead suggestions and surfacing the hierarchy of the controlled vocabulary are particularly efficient at directing searchers to the vocabulary term closest to their areas of interest.



## Collection-Specific Vocabularies Enable Search and Browse Features in Interfaces

Providing type-ahead (sometimes called “predictive search”) suggestions based on the controlled vocabulary used to index/catalog a collection helps to eliminate the guesswork involved in inventing search terms (as illustrated above). As such, it is especially helpful to provide type-ahead suggestions for non-preferred terms (synonyms and other variants) and redirect the searcher to the results for the preferred version of the term.

Surfacing the hierarchy of the controlled vocabulary used to index a collection has several benefits. In addition to allowing the user to browse the vocabulary to find topics of interest, it can also provide a good overview of the breadth of topical areas covered in the collection.

Hierarchy browse also allows the user to explore the depth (granularity) of the vocabulary which, if the vocabulary is well formed, will correspond to the granularity of the collection.

Neither of these benefits can be achieved by using free-text search, author-supplied keywords, or some existing vocabulary not customized for a specific special collection.

## References

- Coates, E. (1988). The Role of Classification in Information Retrieval: Action and Thought in the Contribution of Brian Vickery. *Journal of Documentation*, 44(3), pp.216-225.
- Hedden, H. (2010). Taxonomies and controlled vocabularies best practices for metadata. *Journal of Digital Asset Management*, 6(5), pp.279-284.
- ISKO UK Report, “The Great Debate” discussing the relevance (or irrelevance) of taxonomies in modern information retrieval: <http://www.iskouk.org/content/great-debate#EventReport>
- Jackson, D. (1971). Classification, Relevance, and Information Retrieval. *Advances in Computers*, 11, pp.59-125. doi:10.1016/S0065-2458(08)60630-0
- Jain, Y. and Wadekar, S. (2011). Classification-based Retrieval Methods to Enhance Information Discovery on the Web. *International Journal of Managing Information Technology*, 3(1), pp.33-44.
- Janda, K. (2014). How Do You Say, 'Read Me?' Or Choosing Keywords to Retrieve Information - Social Science Space. [online] Social Science Space. Available at: <https://www.socialsciencespace.com/2014/10/how-do-you-say-read-me-or-choosing-keywords-to-retrieve-information/> [Accessed 20 Sep. 2018].
- Jansen, B. J., & Rieh, S. Y. (2010). The seventeen theoretical constructs of information searching and information retrieval. *Journal of the American Society for Information Science and Technology*, 1517-1534. doi:10.1002/asi.21358
- Mitchell, M. (2017). From Byzantine Art to Winnie-the-Pooh: Classifying a private library. *Catalogue and Index*, (188), 43-45. Retrieved September 1, 2018, from [https://archive.cilip.org.uk/sites/default/files/media/document/2017-10/ci\\_188\\_mitchell\\_classifying\\_private\\_library.pdf](https://archive.cilip.org.uk/sites/default/files/media/document/2017-10/ci_188_mitchell_classifying_private_library.pdf).

I've been a cataloguer for some years but have only just started training to become an indexer with the Society of Indexers. I can now see that there are many parallels between cataloguing and indexing and I am often expanding my knowledge of one activity through the other. The clearest example of a task in which the two areas are intertwined is when I classify theses in our institutional repository. Our current repository platform is EPrints using the Dublin Core Metadata Element Set. This allows us to assign subjects to research outputs so that they are indexed and available to users through access points in our discovery layer (Primo). I'm going to avoid discussion about the systems involved here and their interaction with each other and am going to focus on the details of this task and try to understand some of the benefits and flaws of the current workflow.

When a masters or a doctoral degree is awarded the cataloguing team are notified and we create an entry in the repository and upload the thesis manuscript applying any embargoes where necessary. Once the record has been created we classify the thesis by selecting terms from a subject tree built on Library of Congress authority records and structured using the Library of Congress Classification scheme (LCC). This is also the scheme we use for organising certain sections of our print collections on the shelves.

The top level headings in the subject tree are the main classes which you can see here in the [Library of Congress Classification Outline](#). Our repository software displays a list of these and each heading or 'subject node' can either be selected by clicking 'add' or expanded by clicking '+' to show any further subject headings or 'children'. In this way the tree has a hierarchical structure that nests multiple levels of subheadings under the top level entries potentially allowing a very specific classification to be constructed in much the same way we would when creating a classmark for a print book. If, for example, the research is about Don DeLillo then we could start at P Language and Literature and progressively click through, expanding the headings until arriving at the required level of specificity.

P Language and Literature

PS American Literature,

PS0700 Individual authors,

PS3550 1961-2000,

PS3554.E4425 DeLillo, Don

This process can be repeated and multiple subjects can be selected in order to classify a thesis which spans subject areas. Our discovery layer will display the final term in the string (e.g. DeLillo, Don) minus the class mark, under the 'Subject' facet as an access point in the same way it would display the contents of a 6\*\* field from a MARC record.

If we are unable to find a suitable classification for the thesis then we can add a new subject node as required. For example if we needed to classify our first PhD thesis about Madeline DeFrees we would need to create an entry for her in the tree as a child of the 1961-2000 node. Her subject would display above DeLillo as the LC has assigned her the call number of PS3554.E4 meaning she files before him (but after Charles Bukowski who is at PS3552.U4).

Areas of LCC such as this where author numbers are used lead to small sections within this index where the entries are filed chronologically then alphabetically like this...

PS3500 1900-1960

PS3550 1961-2000

PS3551.U77 Auster, Paul

PS3552.U4 Bukowski, Charles

PS3554.E4 DeFrees, Madeline

PS3554.E4425 DeLillo, Don

PS3563.C337 McCarthy, Cormac

PS3566.Y55 Pynchon, Thomas

PS3568.O855 Roth, Philip

...but in the main it is hierarchically organised and hence displays the bias of the LCC original creators. The structural problems with the hierarchy are embedded but we do have the choice to reject some of the more outdated and questionable LCC terms and codes here. If an entry is offensive or inappropriate or we decide it is not really in current usage then we have the option not to add it to our classification tree. Much like we adapt the LCC scheme to fit our own print book collections and shelving requirements we can adapt it for the repository if we deem it necessary. This is a plus point for an index that grows organically with a collection (as opposed to one built on imported data), we can pick and choose as we go.

Assigning subjects is not an easy task and requires careful analysis of the document. We always read through at least the abstract and try to work out what the thesis is about, in other words we try to understand what the metatopic is. This is a particular challenge for PhD theses because by their nature they contain new concepts that may never have been explored before. We use the LCC schedules and classes from existing items in our own collections to help us.

Although, as I said, we do have a choice about what we add to this index it is still necessarily created from a controlled vocabulary and with that comes some problems. The person doing the indexing has to know how the classification scheme works in order to assign their terms. It takes training and experience to become efficient at this task. Say we had a thesis about pubs, you'd have to be at least a little familiar with LCC to know that you might need to look under the Technology heading in order to find TX0950 Taverns, barrooms, saloons. This has connotations both for the indexer and the user because although the LC authority file might have linked an equivalence relationship between two concepts, this is not translated into our index and synonymous terms would need to be searched for individually by the user.

So why bother with this classification tree, why not have free text entry for subject terms? The repository does have an option for keywords to be entered into the records with the rest of the metadata about the item and you are able to specify them in the advanced search function of the repository itself. However when I tested a few of these it seems they are not indexed in our discovery layer and seeing as the majority of users are accessing our resources from our main Library Search pages we could say that these keywords do not aid discovery as well as terms from the controlled vocabulary of the subject tree. Using the LCC scheme and authority files also allows for disambiguation because qualifiers such as dates are available for otherwise identical entries (authors with the same name for example).

Another reason for assigning subjects from the classification tree is that we are able to classify theses by their theme and area of study. If we were to rely on keyword searches that simply return matches of words from the abstract or title it is not adequate because the metatopic of the thesis may not be mentioned by name. If it is merely implied or buried in specialist terminology then it may need teasing out and highlighting and it is the job of the cataloguer to do this.

Indexing the subjects of theses is a tiny piece in the puzzle of the changing scholarly communications scene and as this scene develops I expect we are going to find systems are better integrated and processes for metadata creation will become more automated. However if there's one thing I have learnt from my indexing training it's that the initial selection of subject terms cannot be satisfactorily automated. Even when we are able to create links and share metadata totally seamlessly, it is unlikely that there will be a way to replace the human intellectual input in the analysis of texts at the academic level. It is through the process of creating new subject nodes described above that we have gradually been able to expand the classification tree over the years, producing ever richer subject options for future cataloguers/indexers to use and add to, hence continually improving the discovery potential for the university's research outputs.

## References

Batley, S., 2005. *Classification in theory and practice*. Oxford: Chandos Publishing.

Library of Congress. 2015. *Classification and Shelflisting*. [ONLINE] Available at: <https://www.loc.gov/aba/cataloging/classification/>. [Accessed 31 July 2018].

Library of Congress. 2016. *Library of Congress Authorities*. [ONLINE] Available at: <https://authorities.loc.gov>. [Accessed 31 July 2018].

Library of Congress. 2015. *Subject and Genre/Form Headings*. [ONLINE] Available at: <https://www.loc.gov/aba/cataloging/subject/>. [Accessed 31 July 2018].

Society of Indexers 2010, 2014 and 2016. *Training in indexing*, 4th edition. Sheffield: Society of Indexers.

### Introduction

Today it is entirely possible to create an index solely using digital devices but this is a very recent development in the centuries-long history of this particular information retrieval tool. This article discusses the mechanisms of creating indexes from ancient to modern times. The focus is on stand-alone books rather than journals, databases or websites but many of the same principles apply to those also.

### Why create an index?

Indexes are tools to enable the content of a work to be accessed in a way other than reading it in entirety; a means to swoop in at a precise point to find certain information. People have always wanted to be able to find, or re-find, a particular topic within a longer piece. Wellisch (1983) says that:

*the root of the word [index] expresses a fundamental communicative action that transcends all languages and probably preceded them in the evolution of man: to show something to another human being by pointing to it with an outstretched forefinger.*

### Who creates an index?

In the very earliest instances of indexes it was often the owner, or at least reader, of the book who created the index, in much the same way that we often create our own mini-indexes to recipe books or family photo albums on our own shelves today. These days it is usually the author, a professional indexer or someone at the publishing house/academic department who is tasked with the job.

The indexer is a shadowy sort, rarely acknowledged and it is not at all uncommon for a professional indexer, on confessing their chosen career, to be met with a delighted '*Oh, I never realised someone actually did that*'. There are few famous indexers and only one known martyr to the cause, William Prynne who had his ears cropped for his all-too-revelatory index to *Histrio-Matrix* in 1634 (Archibald 2014). G. Norman Knight (1968) documents the emergence of the professional indexer in the 18<sup>th</sup> century, describing their '*inferior status – a Grub St Hack*'.

Professional indexers these days, however, are highly trained, often freelance and use a variety of tools, both specialist and generalist, to create their products. Today there are directories of professional indexers and Hazel Bell's book of mini-biographies is enlightening if you want to find out more about the kind of person who takes it up as a job (Bell 2008).

### Underlying analytical skills

A good index points to all the useful information within the text, without being bloated by references to irrelevancies. Therefore, no matter who creates an index, the basic underlying analytical skills are the same regardless of the technicalities of producing it.

The skill of the indexer is in the small, strategic decisions which cumulatively make an effective index.

Archibald (2014) says:

*to do it well requires an ability to quickly synthesize information and conceptualize the thematic relationships within that information [...] A good indexer excels at communicative terseness [and] the golden mean is an indexer who is objective enough to anticipate the needs of a wide variety of readers and subjective enough to parse the central importance of the work*

So first the indexer must identify names and concepts, usually by reading the work cover-to-cover and always with the reader in mind. Secondly they choose a word or words to describe them, which may or may not be identical to the phrasing in the text. Thirdly the indexer must find a way of indicating the destination point; a 'locator'.

### Physical creation of the index

Beare (2007) looks back to the very first pictorial images in cave art and hieroglyphics which include elements of indexing, likewise do the 3,000-year-old *I Ching* (Humphreys 2011) and ancient Hebrew texts (Weinberg 1999). So we know that for probably thousands of years people have considered the labour that goes into producing some form of an index a worthwhile investment in the work's future. Handwritten scrolls, some of the earliest known forms of long texts, did not lend themselves easily to indexing but the codex, which is more book-like in nature, certainly did.

There is debate over when the first subject (rather than name) indexes came to be produced. Ancient Roman Seneca provided Lucilius with a page of notes directing him to certain sections so that he wouldn't have to read the entire work, for example (Archibald 2014). Subject indexes also by definition trace back to the origins of alphabetisation itself (otherwise many early 'indexes' are rather better described in modern terms as 'tables of contents'). This goes back to at least Alexandria in 3rd century BC Egypt when Callimachus created a catalogue for the famous library which was divided into subjects by alphabet (Wellisch 1994). There are alphabetical subject indexes to a collection of quotes and aphorisms called the *Apophthegmata* dating from the 6<sup>th</sup> century (Boardley 2018) and we know that there were many indexes to manuscript books. What we know very little about is the nitty-gritty of *how* these indexes were created, although we can imagine that the process was laborious and mentally taxing. Cornog (1983) notes that there is evidence that such indexes were often handwritten at speed, under time pressures that would be familiar today!

The growth of printing led to a real explosion in the prevalence of indexes although we still know little about the actual compilation practices in early days apart from that it was hard work. Wellisch (1994) talks about one of the very earliest printed indexes, the elaborate and complex *De arte praedicandi* of St Augustine, and says:

*the anonymous editor and indexer says in his preface that he had collated the text of several manuscripts which he had found in different places, though he may have improved on his models because he also claims that it had taken him a long time to compile the index.*

Duncan (forthcoming) quotes a fascinating exchange between John Oldmixon, creator of an index, and his client saying that it '*has cost me a great deal of Pains & richly deserves 121 [coins]*'.

In other words, good indexing takes time and effort, but it is worth it. *De arte praedicandi* is a book on the art of preaching and such a practical book would be relatively useless without a means to look up particular readings or themes. And if it is useless, who will buy it, or recommend it to their fellow preachers?



## *Pages, paragraphs and folios*

The question of what locators to use is interesting. It was only towards the end of the 15<sup>th</sup> century that foliation and pagination became common. Before that indexers used chapters, sections, paragraph numbers, marginal letters and an array of other means to indicate location (Wellisch 1986). In the earliest handwritten indexes the index-creator often had to write in their own page/folio/paragraph numbers – this had the advantage that, just as in the modern digital environment, the index could be copied when the document was reproduced even if the pages were different sizes. Line numbering was another common device.

## *Card-based indexing practices*

Throughout the 17<sup>th</sup> to 20<sup>th</sup> centuries indexes became a key part of scholarship and numbered pages became the *de facto* standard. In the pre-computer age indexers would have read the book, just as we do today, making paper notes as they went along as to concepts and names that needed referencing along with their locators. A common method was to use ‘index cards’, typically 5x4 inch cards with one concept/name per card which could then be reshuffled into an alphabetical order and the index copied up from the cards. Alternative methods did exist, such as thumb-tabbed notebooks (Stallybrass 1968). Jobbing freelancers were booked by letter or phone and hefty packs of paper proofs sent out with the resulting index mailed back along with a paper invoice and the hope of banknotes or a cheque by return. Wheatley (1902) describes the process of creating an index without a computer – cutting up slips of paper, mixing glue, alphabetising them by physically sorting them and hoping that a breeze doesn’t come in and disturb the arranged-but-as-yet-unstuck papers before copying the resulting index by hand or typewriter. Bell (2008) quotes a Margaret Anderson who said that ‘*the main enemies [of the indexer] are hurricanes, housegirls and cocktail parties*’.

Gradually with the advent of computers in the 20<sup>th</sup> century more and more of the physical labour moved to being handled by a computer and today’s enemies are more like broadband freezes, BACS errors and hard-drive failure.<sup>1</sup> Word processors came along first, enabling on-screen editing, spell-checking and even rudimentary A–Z ordering. Specialist indexing software produced a sea-change in how indexers work but also of importance for indexers were the advent of email, pdfs, Office-type software, broadband internet access and then digital publishing processes, tagging and embedding and online workflow processing.

## *Specialist indexing software*

The first specialist indexing software MACREX dates from 1981, a pre-Windows era (Beare 2004). The principles of entering content in a standard form and determining formatting, layout and sorting at the time of output (i.e. to a printer or a word-processing file) are common to all indexing software. There are three main indexing software providers, MACREX, CINDEK and SKY, and they share commonalities as well as having their own specialist features. All of them have developed iteratively in response to user requirements as well as the demands of the publishing industry and provide a variety of useful features such as sorting, formatting, error checking, spell checking and editing functionality.

Publishing software platforms such as InDesign, FrameMaker or LaTeX can have in-built indexing so the indexing is done directly on the live publisher’s file. Microsoft Word also has indexing capabilities although many indexers prefer not to work directly in Word itself but rather to use a specialist intermediary program such as DEXembed, WordEmbed or Index-Manager to provide either an embedded or tagged index.

---

1. Hard to deny that it sounds more fun in Anderson’s time.



Fully-automatic index programs are not yet available, or at least not ones that produce a usable index. However, semi-automatic programs using linguistic and statistical algorithms are becoming available, such as TExtract and PDF Index Generator. These create an initial draft from a pdf proof which can then be formatted, expanded and adjusted before reintegration into the text.

### Indexes embedded or tagged in the text

Up to this point we have looked principally at indexes created as separate documents which are bundled with the text at print stage. A subsequent development, however, is indexes which are embedded in the digital file itself or created using tagging procedures, including XML. This is used both for print books, allowing the index to be reproduced for e.g. hardback and paperback versions which have different page numbers, and for books designed to be used in a digital environment. In this case the index is click-through so there are questions to be debated about how to indicate where there is more than one mention, or the scale of mentions (e.g. the difference between a page range of 10 pages and a single reference). The index entries are linked to the text itself beneath the on-screen words in the form of hidden coding and can therefore be added earlier in the process than the traditional final page-proofs stage. The analytical skills remain the same though – the indexer must still identify what is salient, choose a word or phrase to describe it and then mark out a location to point to. Some publishers have developed their own proprietary tagging systems; others use commercial products.

### Conclusion

Indexes have been around for centuries for the simple reason that they're extremely useful. This brief overview has discussed how even though the basic analytical skills haven't changed, the tools and mechanics of index creation have transformed enormously with exciting new technologies emerging all the time. The future holds more developments in automated indexes, multilingual access, indexes across several publications and the continuing development of digital publishing.

### References

- Archibald, Sasha (2014) 'Indexes, in praise of.' *The Indexer* 32(4), December, 142–148.
- Beare, Geraldine (2004) 'MACREX – a history.' *The Indexer* 24(1), April, 18–20.
- Beare, Geraldine (2007) 'Past, present and future.' *The Indexer* 25(4), October, 257–264.
- Bell, Hazel (2008) *From Flock Beds to Professionalism: A History of Index-Makers*. Hertfordshire: Oak Knoll Press.
- Boardley, John (2018) *Index: A Brief History*. <https://ilovetypography.com/2018/08/24/a-brief-history-of-the-index/> accessed 28th August 2018.
- Cornog, Martha (1983) 'A history of indexing technology.' *The Indexer* 13(3), April, 152–157.
- Duncan, Dennis (forthcoming) *Table of Discontents: A History of the English Book Index*. Oxford: OUP (accessed via Society of Indexers' conference 2017).
- Humphreys, Nancy K. (2011) 'The world's oldest profession: indexing?' *The Indexer* 29(4), December, 161–165.
- Knight, G. Norman (1968) 'Book indexing in Great Britain.' *The Indexer* 6(1), Spring, 14–18.
- Stallybrass, Oliver (1968) 'An unusual method of making a book index.' *The Indexer* 6(1), Spring, 4–13.
- Weinberg, Bella Hass (1999) 'Indexes and religion: reflections on research in the history of indexes.' *The Indexer* 21(3), April, 111–119.
- Wellisch, Hans H. (1983) '"Index": the word, its history, meanings and usages.' *The Indexer* 13(3), April, 147–151.
- Wellisch, Hans H. (1986) 'The oldest printed indexes.' *The Indexer* 15(2), October, 73–82.
- Wellisch, Hans H. (1994) 'Incunabula indexes.' *The Indexer* 19(1), April, 3–12.
- Wheatley, H. (1902) *How to Make an Index*. London: Elliot Stock.

## The set-up

An index can provide an entertaining conclusion to a book in its own right. Some indexes can be unfortunately laughable for entirely the wrong reasons, but many authors and indexers choose to add deliberate humour to this final part of the book, and some do this very well. Over the last couple of years, I have been writing a series of blog posts on the comedy book index on my website ([baindex.org](http://baindex.org)), some of which have become articles for *The Indexer* journal (Bain 2017, 2018), continuing the fine tradition of Hazel Bell's reviews of comedic indexes (see Bell 2001a, 2001b). I'm pleased to present an overview of some recent examples for this issue of *Catalogue and Index*.

I have always had a great interest in comedy in my leisure time, and I have indexed a lot of books in my work hours, having been a full-time freelancer since 2001. It's been a great delight to me, as both a comedy and indexing fan, to see that many of my favourite comedy writers have recently produced books that have plenty of humorous material in their indexes. These include Francis Wheen, Charlie Brooker, Richard Ayoade, and the writers of the Alan Partridge and Steven Toast characters, all of which I review below.

## Strange indexes indeed: Francis Wheen

The earliest of these were two books by Francis Wheen, journalist and deputy editor of *Private Eye* magazine. *How Mumbo-Jumbo Conquered the World* was published in 2004 and *Strange Days Indeed* followed in 2009. The books provide commentary on the political and social changes, and associated delusions and paranoia, of the late 20th and early 21st centuries. Both books were indexed by Francis Wheen himself, who proves to be a fine author-indexer. The humour in these indexes largely comes from the subheadings used to qualify people in the index. Often these are politicians, from the 1970s in *Strange Days*:

Heath, Edward: as Empress of Blandings, 38  
Nixon, President Richard M.: ... bugs himself, 34–5, 37–9  
Wilson, Harold: ... as big fat spider, 9, 269, 270

And from more recent times in *Mumbo-Jumbo*:

Blair, Tony: ... claims descent from Abraham, 165  
Clinton, Bill: empathy junkie, 194  
Thatcher, Margaret: ... enjoys 'electric baths', 129

This index includes a few interesting subheadings for royalty:

Elizabeth II, Queen: accused of cocaine-smuggling, 149  
Philip, Prince: enjoys *Flying Saucer Review*, 136; praised by extra-terrestrials, 137–8

It also has some nicely balanced wording to show the comeuppance of certain figures:

Aitken, Jonathan: admires risk-takers, 59; goes to jail, 60  
Ronson, Gerald: eulogised by Jonathan Aitken and Jeffrey Robinson, 59; goes to jail, 60

There is also an extensive entry for 'God' in the *Mumbo-Jumbo* index in the UK edition, including subentries for various things that people had claimed as God's work, for example '*arrives in America; helps vacuum-cleaner saleswoman; produces first self-help manual*'. Interestingly, this index was scrapped and redone for the US market. Francis Wheen told me via Twitter that his US publisher advised him, '*Over here, we don't like jokes in indexes.*' What a shame. In my view, the indexes just enhance the humour of the books.

### **Charlie Brooker's *Burn*, *Dumb*, *Hell* and *Hate***

The same could be said about the indexes of writer and presenter Charlie Brooker, namely the ones for *Screen Burn* (2005), *Dawn of the Dumb* (2007), *The Hell of It All* (2009) and *I Can Make You Hate* (2013). These books are collections of his *Guardian* newspaper columns. He confirmed for me on Twitter that his publisher Faber & Faber supplied him with the indexes in basic form, which he then 'joke-i-fied'. In the *Burn* and *Dumb* indexes, most of the humour comes from amusing subentry wording relating to politicians and TV stars. Several are also rather sweary but here are a few tamer examples from *Burn*:

Barker, Linda: inferior to mop, 252  
Brown, Gordon: lesser twerp than Blair, 285  
Davidson, Jim: provokes tooth-grinding, 7

And from *Dumb*:

Cameron, David, justifiable abuse of, 275–6  
Edmonds, Noel, exploring dual realities, 139–40  
Hewer, Nick, face like an eagle peering at a banknote, 155

By the time of the *Hell* and *Hate* indexes, Brooker is having even more fun with the 'joke-i-fying'. In the *Hate* index, Brooker deliberately misnames many of the figures referred to (correctly) in the text:

Blair, Tiny, 100, 133  
Chiles, Atrium, 257  
Clegg, Knock, 99, 100, 101  
Van Outen, The Niece, 51

There is also a nice sequence of circular cross-references in *Hate*:

ceaseless repetition, *see* déjà vu  
déjà vu, *see* make it stop  
endless loop, *see* unbreakable cycle  
I can't, I just can't, *see* endless loop  
make it stop, *see* I can't, I just can't  
unbreakable cycle, *see* ceaseless repetition

Both the *Hell* and *Hate* indexes display a playful sense of toying with the index reader. From *Hell*, we have:

letters of the alphabet, variously configured, 1–388  
lying, 18–19, 131–3, 225–6, 299, 346–8 (one of these is a lie)  
numbered pages, 1–388  
random page numbers, 8, 44–9, 70, 84, 213–28, 337

In *Hate*, there are many more:

- arbitrary choice of page, 14
- scanning the page for a reference that isn't there, 67
- section of air behind book, 410–495
- time: regrets about wasting, like you're doing now, 29

Some very odd things are going on elsewhere in this index:

Someone is behind you; not physically, but in spirit; call it a presence if that helps. This isn't an index entry; it's a warning: someone is standing directly behind you *right now*

And, spookily, near the end:

you are imagining these words in the index

### **Back of the book/net: Alan Partridge (aha)**

All the books discussed so far are non-fictional works about real-life events and people. Some other recent comedy books are fictional works presented in the form of memoirs. As they are structured as autobiographical texts, they also include indexes like standard works of non-fiction.

Two excellent examples are the Alan Partridge parody memoirs *I, Partridge: We Need to Talk About Alan* (2011) and *Alan Partridge: Nomad* (2016). Alan Partridge is an established comedy character, played by Steve Coogan since the early 1990s. The books are written by Neil Gibbons, Rob Gibbons and Steve Coogan (plus Armando Iannucci for *I, Partridge*). *I, Partridge* is presented as the autobiography of Alan Partridge, with *Nomad* as its 'beautifully punctuated follow-up', describing Partridge's mission to walk a 160-mile journey from the fine city of Norwich. As with a standard (auto)biography index, the largest entry in *I, Partridge* is for the main subject himself, with richly detailed subheadings and sub-subheadings:

- Partridge, Alan Gordon
  - dress and grooming
    - aftershave (Pagan Man) 122
    - dressing gown, when alone 132
  - food and drink
    - Interfered-with sandwiches 24
    - Toblerone addiction 203–4, 213–23, 228–9, 231
  - homes and dwellings
    - static caravan 254–5
    - 'unstatic' caravan 260

In *Nomad*, the main entry for Partridge is considerably less detailed, and not placed where you might expect to find it:

- I (Alan Partridge) 1–285
- Partridge, Alan – see I

The entries for Partridge's parents in *I, Partridge* offer intriguing insight:

- Partridge, Dorothy (mother)
  - compared to Rover 800 177
  - neither nice nor important 141
- Partridge, Snr (father)
  - butterfly tennis 7
  - death 244, 247
  - Location post-death 249

As does the one for his ex-wife:

- Partridge, Carol *née* Parry (ex-wife)
  - body shape, attempted 260
  - dog-like hair 37
  - Jealous of Sue Cook 71

The *Nomad* index is much shorter than the first one but still offers several nuggets:

- Cherubs
  - physical strength 231
  - aeronautic capability 232
- Countryfile*, that woman who sued 10
- Haddaway, a man that looked like 257
- Jam bombs 149
- Jambon 149
- Last Post, the
  - whistled while miming a bugle 36
  - actually bugled, albeit badly 213
- Phalanx, nice use of the word 73

Neil Gibbons told me via Twitter that for the *Nomad* index, 'someone *at the publishers did a basic one and we mucked about with it*'. I was also pleased to be contacted by Mark Bolland at HarperCollins, who confirmed that he was the original indexer for *I, Partridge*. He kindly allowed me a look at his original, unexpurgated index file, which was considerably longer and even more Partridgean, but unfortunately much of this was cut before publication. What remains is still great fun though.

## Time for Toast

Similar humour can be seen in the index to the spoof 'part memoir, part "how to act" manual' *Toast on Toast* (2015), based on the life of 'Steven Toast', renowned thespian played by Matt Berry in the TV series *Toast of London*. The book was written by Matt Berry and Arthur Mathews, and Arthur Mathews confirmed for me that he wrote its index. Again, the subheadings for the main character give a measure of Toast's self-regard:

- Toast, Steven: early signs of talent, 16; healthy sex drive, 67, 89, 99, 200; hugely successful voiceover career, 89, 90; receives ecstatic reviews, 78, 99, 100–06; success with women, 80, 90, 110, 111 ...

Toast's main acting rival, Ray Purchase, is afforded a detailed entry in the index:

- Purchase, Ray: lack of talent, 56, 67, 89–98; incompetence as actor, 67, 88, 99, 122, 134; ... peculiar appearance, 88, 90; ... senseless cruelty towards animals, 67; kills Michael Ball, 91

As is his wife:

Purchase, Mrs, 56, 78; failed and loveless marriage to Ray, 89, 90–9; Toast has sex with, 67, 68, 70–2; ... admiration of and loyalty to Toast, 88; takes part in drone strikes for US government, 81; appearance on *Strictly Come Dancing*, 188

There are some gloriously named fellow (fictional) thespians:

Aaaaaadams, Aaaaaalan, 98  
Fandango, Clem, 78–89  
Frumpty, Mews, 67, 68, 80, 82, 84, 86  
Fufoon, Besus, 89  
Residue, Shorley, 78, 89, 99, 100  
Spraymount, Billy, 1  
Strepsils, Henty, 88

Certain things are in the index for comedic value alone and do not relate to the text at all:

4, Channel, 4  
101, Dalmations, 101; Toast counts individually, 102  
Ali, Muhammad, 7; floats like a butterfly, 88; stings like a bee, 603

It's all rather weird and wonderful.

## Ayoade on Ayoade

Speaking of index oddities, there is also much to be enjoyed in the book indexes of Richard Ayoade, comedy actor/writer and film director. His two books *Ayoade on Ayoade: A Cinematic Odyssey* (2014) and *The Grip of Film* (2017) are, as the titles indicate, based in the world of cinema. I have not yet found out who indexed these but I suspect it may be similar to the Charlie Brooker situation, with a basic index 'joke-i-fied' by the author. *Ayoade on Ayoade* is a parody of Faber & Faber's *Directors on Directors* series, with one 'Ayoade' persona as interviewer questioning another 'Ayoade' as film director. The alternative persona he adopts in *The Grip of Film* is someone else entirely: a foul-mouthed director named Gordy LaSure. The index to *The Grip of Film* is lengthy, often hilarious, and even comes with an introductory note denoting various typographical cues for different kinds of entries. Much of the content of this index however is extremely sweary, reflecting the '*relentlessness of Gordy's profanity*', so I'll say no more about it here, but a few choice examples are included on my blog.

The *Ayoade on Ayoade* index is much briefer but equally peculiar. Given that this is supposed to be a film director's musings, a reader might expect an index full of directors and films. There are a few but not the ones you might predict (examples include such classics as *Herbie Goes Bananas*, *Scary Movie* and *Speed 2: Cruise Control*). What does appear are various food stuffs:

grapes, squelchy, 260  
guavas, bombardment of, 267  
nectarines, pelting with, 267  
puddings (custard-y one and a lighter one), 190  
Shakey Jake (milkshake), 3

And most of all just many, many strange things:

- alarms, egg-shaped panic, 122
- dew, mature stag glistening with, 255
- guff, inexhaustible tub of, 126
- mechanically closing door, power-slide under a, 249
- pescetarian, bitter feud with a, 254
- sheep, ten billion gently gavotting, 249

All of these are indeed mentioned on the pages given, but they are not what any reader would think to look up, nor what these pages are actually *about*, but then that's not really the point.

### Back (of the book) to the future

A forthcoming blog post will be on *Their Brilliant Careers* by Ryan O'Neill (2018). This is a wonderfully funny book which includes several additional gems in the index. The book purports to be a collection of profiles of 'sixteen extraordinary Australian writers' but the whole thing is a spoof, including the exaggerated author persona of 'Ryan O'Neill' here assumed by the real O'Neill. The reader is told halfway through the book that its index was written by Rachel Deverall, the estranged and now deceased wife of 'Ryan O'Neill'. Deverall is also one of the writers profiled in the book, herself being a made-up figure. The index reveals Deverall's attitude on her husband and his book:

- barefaced lies, 121–37
- delusions of grandeur, 259
- egomaniac, *see* O'Neill, Ryan
- hackwork, 28, 37–9, 66, 95–9
- mistakes, 1–259
- mistress, *see* Zoellner, Anne
- pyromaniac, *see* O'Neill, Ryan

There are other much worse (and funnier) cross-references to O'Neill in the index.

As regards my own indexing work, I try to include a few entertaining entries when the tone of the book warrants it. This summer, I was lucky enough to index *Soupy Twists! The Full, Official Story of the Sophisticated Silliness of Stephen Fry and Hugh Laurie* by Jem Roberts (2018). Jem and I thought we'd have fun with a few of the index headings, reflecting the daftness of the comedy source material. Time will tell how many of these entries made the publisher's final cut, but if enough of them make it to print (publication due early September), I will blog about this one too.

### The punchline

In this whistle-stop overview, I hope I have given some flavour of the kinds of humorous indexes that are currently being published by some of my favourite comedy writers. There is much more in each worth exploring. All of the indexes still work as perfectly usable information retrieval devices, aside from things included purely for comedic value, and in my eyes they provide considerable added entertainment value at the backs of their respective books. I'm always on the lookout for new examples so I would be keen to hear of any further suggestions. May there be many more laughable indexes.



## References

- Ayoade, Richard (2014) *Ayoade on Ayoade: A Cinematic Odyssey*. London: Faber & Faber.
- Ayoade, Richard (2017) *Richard Ayoade Presents: The Grip of Film by Gordy LaSure*. London: Faber & Faber.
- Bain, Paula Clarke (2017) 'Back of the book, back of the net: the comedy book indexes of Partridge and Toast.' *The Indexer* 35(1), 18–24.
- Bain, Paula Clarke (2018) 'Strange indexes indeed: the wit of Francis Wheen as author–indexer.' *The Indexer* 36(2), 42–48.
- Bell, Hazel K. (2001a) 'An index for Thalia.' *The Indexer* 22(3), 147–148.
- Bell, Hazel K. (2001b) *Indexers and Indexes in Fact and Fiction*. London: The British Library.
- Brooker, Charlie (2005) *Screen Burn*. London: Faber & Faber.
- Brooker, Charlie (2007) *Dawn of the Dumb*. London: Faber & Faber.
- Brooker, Charlie (2009) *The Hell of It All*. London: Faber & Faber.
- Brooker, Charlie (2013) *I Can Make You Hate*. London: Faber & Faber.
- O'Neill, Ryan (2018) *Their Brilliant Careers: The Fantastic Lives of Sixteen Extraordinary Australian Writers*. Rickmansworth: Lightning Books.
- 'Partridge, Alan' [Rob Gibbons, Neil Gibbons, Armando Iannucci and Steve Coogan] (2011) *I, Partridge: We Need to Talk About Alan*. London: HarperCollins.
- 'Partridge, Alan' [Neil Gibbons, Rob Gibbons and Steve Coogan] (2016) *Alan Partridge: Nomad*. London: Trapeze/Orion.
- Roberts, Jem (2018) *Soupy Twists! The Full, Official Story of the Sophisticated Silliness of Stephen Fry and Hugh Laurie*. London: Unbound.
- 'Toast, Steven' [Matt Berry and Arthur Mathews] (2015) *Toast on Toast: Cautionary Tales and Candid Advice*. Edinburgh: Canongate.
- Wheen, Francis (2004) *How Mumbo-Jumbo Conquered the World: A Short History of Modern Delusions*. London: Fourth Estate.
- Wheen, Francis (2009) *Strange Days Indeed: The Golden Age of Paranoia*. London: Fourth Estate.

**Paula Clarke Bain** is an Advanced Professional Member and current Marketing Director of the Society of Indexers and has worked on nearly 800 books since going freelance in 2001. She likes comedy, books and indexes and really likes comedy book indexes. She can be found on Twitter as @PC\_Bain and her website is [www.baindex.org](http://www.baindex.org)

## The foundation of the Society of Indexers

Rosemary Dear, Member of Executive Board, Society of Indexers

In 1956 G. Norman Knight (1891–1978), a freelance indexer of 30 years standing, knew of no-one else in that line of work. He had taken up indexing as a part-time activity in 1925, in addition to working as a civil servant and writing innumerable letters, articles and books on freemasonry and chess. He placed an advertisement in the *Times Literary Supplement* on 28 December requesting anyone who was interested in forming a society for indexers to write to him. He received several dozen replies and then hosted 'several discreet luncheons' at the Civil Service Club, of which he was a member, to sound out some of those who had replied with a view to providing an initial complement of officers for the Society. The Society of Indexers (SI) was formally inaugurated at 10.30 a.m. on Saturday 30 March 1957, eighty years after Henry Benjamin Wheatley had founded the first professional association of indexers, the Index Society. The Index Society had merged with the British Record Society a few years later and a representative of this society was amongst the sixty-five people present at SI's inaugural meeting, as well as representatives of Aslib and the Library Association. One of the founder members, Professor William Hechscher, even flew in from Utrecht expressly for this event.

At first the Society was run from the homes of various officers, but as it grew in size it moved out to offices, first in Mermaid Court near London Bridge which it shared with SFEP (Society of Freelance Editors and Proofreaders) and later to a series of locations in Sheffield. Our current office is still in Sheffield in the Woodbourn Business Centre, and run by our full-time office manager. At first the Society was run by a Council, but as the membership increased this was replaced, in 2004, by a smaller Executive Board based around directorates and a larger Consultative Committee which represented the grass-roots membership. At the same time more of the administrative work was undertaken by full- and part-time staff in the Sheffield office. In 2007 the Society celebrated its 50th anniversary and became a company limited by guarantee, which meant that the Executive Board became a Board of Directors which had to run the Society in accordance with the Companies Act. The Consultative Committee continued to meet until 2010, when it was disbanded.

One of the main aims of the Society is to promote excellence in indexing. To achieve this it set up a training course, the first of these being a series of six lectures in 1958 in London. A correspondence course run by the Rapid Results College was established in 1973, followed ten years later by a Book Indexing Postal Tutorials (BIPT) course. The first edition of the Society's distance-learning course, *Training in Indexing*, was published in 1988 and has been updated several times, with the online version launching in 2010. Workshops have been held in London, at regional group meetings and as part of SI's (now annual) conferences, and have also been available online since 2013.

Over the years SI has produced a number of publications. The most important of these is *The Indexer*, a learned journal which first appeared in 1958 as a biannual publication, moving to quarterly publication in 2008. From 2019 it will be produced by Liverpool University Press. The Society's newsletter started life as a typed sheet of A4: it was renamed *SI delights* in 1996 and has been issued four times a year since then, moving to electronic delivery in 2015. *Indexers Available*, a list of qualified indexers designed to help editors find an indexer, is another publication that went electronic. It was first published as hard copy in 1982 and then became available online in 1998 when the SI website was established. In 2004 the hard copy version was discontinued and the online directory was relaunched in 2017 as the Directory of Professional Indexers when the updated website went live.

Awards are also important to the Society. The first award to be created was the Wheatley Medal for an outstanding index, established jointly by the Library Association and SI in 1960. The Carey Award (for services to indexing) was established in 1977, followed by the Bernard Levin Award (for services to SI) in 2000 and the Betty Moys Prize (for the best newly accredited indexer) in 2002.

The Society has had to keep pace with the breakneck speed of technological advances, with its members moving from indexes composed of entries on hand-written slips of paper delivered to compositors to indexes written using dedicated and sophisticated software delivered by email either as free-standing Word documents or in a variety of embedded forms. A newsletter entitled *Microindexer* was published between 1983 and 1994 to help ease the transition from keeping slips in shoe boxes to computer files. The latest innovation has been the creation of a National Indexing Day, first held on 30 March 2017 to mark the Society's diamond jubilee.

As an indexer, I provide directions for readers and researchers so they can find information. Sometimes I provide directions to material they know they are looking for. Sometimes I provide directions to material that they do not know they are looking for. The challenge in writing indexes for books, monographs, cumulative works, and other closed sets of materials is to provide a good map regardless of what your readers are looking for. The goal of a quality index is simple: to provide a navigation aid to the information.

There is no one clear, best sign as to what direction indexing is going, however. I considered providing a photograph of some of the more challenging road signs I have seen to illustrate this quandary. Recently, while in San Diego, California, I saw a wonderful sign that included labels to “Jaws, 2,535 mi. à,” “Tacos ß,” “Malibu, 150 mi. à,” and other destinations such as “Pipeline.”



*Beachside sign in front of the Hotel del Coronado, San Diego, California (photo by the author)*

The future of back-of-the-book style indexes is full of as many possible destinations, and more. Some of these possible destinations may pique our professional and intellectual curiosity, some may be frightening. Some of these possible destinations may have known and measurable paths to them, some may not. While I cannot say for sure where we will end up, I do have some ideas.

## The path to indexing

First, let me share how I, personally, got here and the path I took. Like many professional indexers (and other professionals), my career path has been serendipitous and unplanned.

One of my first jobs, in the summer of 1979, included filing articles for the Middle East Bureau Chief of *The Washington Post*. At the time, I did not realize the training this exposure to a catalogue system provided me. I found the work and material intriguing, interesting, and satisfying.

Four years later, as a Liberal Arts college student in 1983, I began to work part-time as a typist doing computer data entry. I swore, when I learned to type as a pre-teen, that I would never take a job where typing was all I did. I didn't mind the occasional jobs for classmates and other students who paid me to type their papers and essays. But for this part-time job, I also paid attention to what I was typing.

I worked for a professional, freelance, full-time book indexer: Maria Coughlin. She marked up paper page proofs of books ready for publication, and I then entered index text accordingly into her computer, along with the number of the page her mark-up appeared on. In the beginning, I entered this text into word-processing software.

(Other indexers at the time wrote their index entries for a book on index cards, and then turned that box of index cards over to a typist for conversion into an index document.)

Soon, though, professional indexing software became available and I entered text directly into the computer via one such tool: Cindex™ (available from <https://www.indexres.com/>). This professional indexing software provides automated alphabetization and all sorts of formatting and editing. I was a fast typist and Maria and I worked well together. Over time, I noticed patterns, and sometimes I noticed inconsistencies, or what I thought were inconsistencies. I asked questions about what I was typing. I learned to anticipate how she might mark up text. Our relationship became a de facto apprenticeship.

I continued this work until I graduated from college. At one point, while studying in another state one year, I also worked as a technical writer. When I returned to the area for visits with family, I would again do data entry work and help Maria with indexing. In early 1989, after a couple years of graduate study in mathematics, I again returned to the area and officially joined her company, M. L. Coughlin Editorial Services, as an Associate Indexer. At this point, I also began to mark page proofs myself.

Clearly, I enjoyed the work and it was a good fit. Periodically, other students also joined the company. Some of these others also found the work a good fit, some did not. Even though most who first come to the work are self-selecting, indexing work is not for everyone.

In 1990, I launched my own indexing business: Wyman Indexing. Like Maria, I had a computer, and I worked from a home office. I also continued to work part-time for Maria, at her home office, until 1993.

## The path indexing has taken since

Intellectually, the process of indexing and writing an index is always the same: read, review and analyze (ask and determine, "What is this about?"), record succinctly, repeat until the entire text has been processed, review and edit the index, stop.

This process is not that different for database indexers, though they generally do not have as much control of the final index or database.

Since launching my business, however, indexing work has changed in several ways, just as technology and publishing have also evolved. While I still work from a home office on a computer and with professional indexing software, software tools have changed, the materials I use have changed, the materials I index have changed, and the work process has changed.

I still use Cindex™. In fact, I swear by it. But it has changed and updated itself, continuously. Additional indexing software tools are also available, and I now counsel beginning indexers to test them all before committing to any. I also use Index-Manager (<http://index-manager.net/en/home/>) now, for projects that require embedded indexing.

I no longer work at a desktop computer. Like many of you reading this, my tools have shrunk as technology has advanced. My main computer is a tablet laptop, and I back-up to the cloud. I also rarely use a printer anymore as I email files to my clients (vs printing out manuscripts and sending them via FedEx along with a disk copy of the index file).

I still index books, as well as other materials such as collections of documents, but I used to also index many journals. That is, I used to write cumulative volume indexes for journals, as you would see in the last issue of an annual volume of a periodical. With ubiquitous Internet access, most journals now provide digital copies and full-text search of their materials, and they no longer include so many indexes.

My work process has also changed. I no longer always include page numbers when I type or create index entries. As technology has advanced, publishing has also changed. Instead of working from paper page proofs, I now work from electronic page proofs or from manuscript files. I may be reading from another monitor now, vs a stack of paper as I did when I worked for Maria. When I work from electronic page proofs, I may include page numbers with my index entries, or I may include some sort of unique identifier (UID) for the text I'm indexing. When I work from manuscript files, I may include some sort of UID for the text I'm indexing, or I may use Index-Manager and write the index directly into the manuscript where it is hidden and embedded for generation with final page numbers when the manuscript is ready for publication.

Embedded indexes and UIDs allow publishers to speed up the production process, and make it easier for publishers to reuse content – for new editions including in different formats, such as digital or e-books.

There have been repeated technical challenges as technology has evolved. Some clients have experimented with XML, for example, while they explored publishing editions on the web or for handheld devices or for mobile apps. Other clients have explored ways to embed their indexes. I have had to stay current and on top of technology and all these developments.

There have been repeated fiscal challenges in the face of global competition. Some clients have left me for cheaper indexers and lower production costs in India or other off-shore markets. Most have come back, but not all. I have had some fallow periods, briefly, and am not always booked as far in advance as I may like.

There have been repeated burnout challenges, as well. Generally, I have found there is a lot of work to be done, so much so that I often think I could easily work every day of the week, every week of the year. At one point, when facing burnout, I took a job part-time as a librarian at my children's school. This gave me a change of pace and allowed me to better structure my time. This also forced me to say, "No" to a lot of projects. I have to say, "No" to a lot of projects because I can only accommodate so much work. Some clients have learned that advance notice is best, but many authors and others are unable, it seems, to plan ahead. Some aspects of publishing never change.

## Possible paths forward for indexing

Again, it has been a challenge to stay current with technology. For me, this has been a challenge I have relished. I was an early computer user and studied computer programming as a teenager, and I have enjoyed staying current with technology and computers. I also know there is more I could learn, however, always. I am grateful for new software such as Index-Manager which allows me to keep indexing in a way that is useful for publishers and other clients. Anybody who is considering indexing or a related field should anticipate continuous professional development and education to stay current with technology.

Embedded indexes allow indexes to stay with the books they supplement regardless of how they are being published, and for indexes to be produced before books are in final form. Thus, it seems to me that embedded indexing may be here to stay. I also know of many other indexers who embed their indexes or provide embedded indexing services but who work in Word or InDesign or other publishing tools directly.

But I'm not sure.

Perhaps some publishers are simply leery of letting go of their files, and do not trust embedding not to introduce errors or provide other complications to their workflows.

I, personally, still have more work with UIDs or page numbers than I do embedding.

Students prefer printed textbooks. Many of us still prefer hard copy for much of our reading. Print books are still here. This means that job security for indexers is not in doubt, for now.

I do recommend indexing as a career option. I'm not sure what form indexes will take, or how indexing and the writing of indexes will proceed or with what tools, but indexes themselves do not seem to be going away.

Information access is still key. It's still true that if readers or researchers can't find something, it might as well not exist.

I was grateful for that sign on the beach. Who doesn't appreciate or want to know where to find tacos, or where the sharks are? We all do. It's the same with indexes. We need and want to know where to find information and things, and we appreciate those who point us in the right direction.



## The Problem

In July of this year, I saw cataloguers on Twitter drawing attention to a recent LCSH revision which changed 'Mental retardation' to 'Intellectual disability'. Twitter user violetbfox wrote: "*Ask your library's cataloging folk if these changes must be made manually; if so, ask them to make it a priority.*"<sup>1</sup> I found her call to action inspiring – another of the many ways in which library workers are willing to share experiences and knowledge to help each other improve our services.

Around the same time, I discovered that my (now former) institution's catalogue still had subject headings containing the word 'Negro'.

Even knowing how slow LCSH were to change, I couldn't believe this was still an authorised heading – and of course, it isn't. It hasn't been since 1975.<sup>2</sup>

So what was it still doing in our catalogue?

A brief survey of other academic and research libraries in the UK revealed that this was a widespread issue. Using institution's OPACs and resource discovery tools, I conducted searches for subject headings using these outdated terms. I carried out these searches for over 50 institutions from across the UK. The results were striking. 83% of catalogues surveyed contained at least one subject heading containing the word 'Negro'.

In the majority of cases, the number of records affected was small and confined to material dating from the 1970s and before. These subject headings had been created at a time when these terms were the authorized ones, and had missed out on later updates.

Naturally there are good reasons why local subject headings are not always up to date with LCSH. Time and resources are always limited, and when headings are subject to numerous small changes over time, the benefit of constantly updating your catalogue may well not merit the effort. Ensuring that every instance of 'Cookery, French' instead reads 'Cooking, French' is hardly a high-priority. Limited resources could be better spent on many other things.

But not all heading changes are equal.

We would all hope that no one is actively searching our catalogue using the word 'Negro', but as long as subject headings containing this word remain in our records, our users may encounter them.

---

1. Violet Fox (@violetbfox). "Important #LCSH change alert, from 'Mental retardation' to 'Intellectual disability'. Ask your library's cataloging folk if these changes must be made manually; if so, ask them to make it a priority. #critcat #critlib". 8:33 AM, 18 July 2018. Tweet.

2. Steven A. Knowlton (2005), 'Three Decades Since *Prejudices and Antipathies*: A Study of Changes in the Library of Congress Subject Headings', *Cataloging & Classification Quarterly*, 40:2, 130.

In a recent episode of Turbitt & Duck, Alissa McCulloch made the incisive observation that for many users, the online catalogue may be the first part of the library that they access.<sup>3</sup> The metadata in our catalogues, we know full well, is central to resource discovery. Users go to OPACs to learn what resources they can access, and cataloguers aim to create metadata that reveals the nature, creators, and subjects of these resources. We take care to create accurate metadata because we know that errors or misleading information will hinder our users' ability to find what they are looking for. What about the impact on users of out-of-date subject headings?

In the cases of many LCSH updates, it may be a minor inconvenience only. For example, in a catalogue which indexes the old heading 'Organ – History' as well as the current heading 'Organ (Musical instrument) – History', interested users are likely to find both headings within a short period of browsing the index, since they will sit in alphabetical proximity. A keyword search for 'organ history' will return records with both headings. Users may guess when observing the split headings that one is an updated version, but even users with no contextual awareness of how subject headings are created and structured are unlikely to suffer anything more than an annoyance.

The consequences of out-of-date subject headings like 'Negroes – Great Britain' or 'American literature – Negro authors' are likely to be much worse, in at least two respects.

### The Impact on Users

Firstly, we must consider the potential effect on users, especially users who are Black.

Archaic and offensive terms in book titles will likely be understood as relics of their time. Subject headings, however, represent the voice of the library. Subject headings are our authoritative statements concerning what the book is about. As Olsen puts it, "*LCSH shapes the meaning that is conveyed from a document to a user.*"<sup>4</sup> As libraries, we position ourselves between resources and readers, in a role of facilitation and illumination. When we catalogue and classify and index, we say to our users: 'this is a resource about x and y'. Our catalogue speaks for us, and so we must be mindful of what it is saying.

The presence of 'Negro' and 'Negroes' in subject headings risks alienating and offending our users, especially our Black users. As Biswas argues in a critique of the LCSH 'East Indians', "*A rejection of outdated terminology is central to providing any culturally sensitive tool for resource organisation.*"<sup>5</sup> There are many headings like 'East Indians' which LC still needs to be convinced to change, but 'Negroes' has long since been updated. Excising these subject headings from our catalogues is many decades overdue.

Our OPACs are one of the public faces of our libraries. If we want our users to feel welcome, comfortable, and respected in our library spaces, including our digital spaces, then our catalogues cannot contain these terms.

---

3. Alissa McCulloch, *Turbitt & Duck: The Library Podcast*, Episode 18. Podcast audio, 9<sup>th</sup> July 2018. 17:13-19:17.

4. Hope A. Olson (2000), 'Difference, Culture and Change: The Untapped Potential of LCSH', *Cataloging & Classification Quarterly*, 29:1-2, 66.

5. Paromita Biswas (2018), 'Rooted in the Past: Use of "East Indians" in Library of Congress Subject Headings', *Cataloging & Classification Quarterly*, 56:1, 1.

## The Impact on Access

Secondly, we should consider the impact on resource access.

Resources on particular subjects being split between multiple subject headings disproportionately inconveniences users looking for books in these areas.

LC first changed 'Negroes' to 'Afro-Americans' and 'Blacks', and later updated the former heading to 'African Americans.' If a catalogue contains records using all these different headings, and no effort has been made to update them, this is a barrier to users successfully discovering our collections. A user interested in African American art would have to locate and browse 'Negro art', 'Afro-American art', 'African American art', and 'Black art' to find all resources on the subject. Keyword searches using current terms would be unlikely to retrieve records containing only the older headings. At best, this means the search is more time-consuming and more inconvenient for our user. At worst, this could mean the user does not discover a portion of the resources available.

The unintended (but real) consequence of this is that we create better access to some areas of our collection than others, and this better access has the potential to support existing power hierarchies. Works about white artists – or, in LCSH terms, just 'Artists' – will be easier to locate than works about Black artists.

When we make it harder for users to find material relating to historically marginalised groups, we maintain barriers to learning and research which relates to these groups.

## Call to Action

The presence of certain out-of-date LCSH in our catalogues has a detrimental impact both on our users and on the learning and research our collections are used for.

We as libraries are responsible for critically analysing the statements we make, and the statements our catalogue makes for us. Language is continually evolving, and it is likely that various words that are considered appropriate now will be contested in the future.<sup>6</sup> This doesn't mean, however, that changing our language now shouldn't be a priority.

We may not have the resources to ensure our subject headings are perfectly up-to-date, but we should be aware of the particular outdated headings which are more harmful. We can act to correct those, at least, and we can do that now.

---

6. Emily Drabinski (2008), 'Teaching the radical catalog' in *Radical Cataloguing: essays at the front*, edited by K.R. Roberto, p. 202.

### Abstract

The Language Centre library has a collection of online external resources on its website for the languages it also has in the physical library. The external weblinks were presented alphabetically by language. The website was moved to a more professional platform, Oxford Mosaic. The web architecture was already made and it led to some rethinking of the online classification as it moved from alphabetical to partly alphabetical and geographical. Some other language classifications are being discussed as well as a study of other language centres and libraries with substantial foreign languages collections.

[Oxford University Language Centre library](#), like many language centres, possesses an [online library](#) as well as a physical one. While online resources created by our Language Tutors, or resources using licences are on a VLE (Virtual Learning Environment), the website also offers a collection of external links for learning languages and some cultural websites. These links are mostly external and free to use, for example grammar exercises quizzes, or resource gateways, mostly from universities all around the world or, simply, links to museums and galleries.

This is an example with an extract of our [Italian page](#):

#### Grammar

- [Centro Studi Italiani eclass](#) Grammar fact sheets and exercises
- [Dig-Itale Eduware](#) exercises from J.A. Campana at the University of Toronto
- [Impariamo l'Italiano](#) check your grammar and learn Italian sayings

#### Vocabulary

- [Smart Phrase, Smart Phrase Online Phrasebook](#) Useful vocabulary and phrase finders in Italian.

#### Italian language resources

#### Culture

##### General

- [Biblioteca digitale IntraText Online](#) digital archive of Italian literary, theatrical and religious works
- [City of Florence](#) Official web site of the Municipality of Florence
- [Global Arte](#) Dedicated to history of art, from its origins to now
- [Italian Cultural Institute](#)
- [Letteratura.it](#)
- [Literature page from Zanichelli](#) excellent documentation on many Italian authors, there is [also this page available](#) and [tests available too](#)
- [Italian woman writers](#) from the University of Chicago
- [Ministero per i Beni e le Attività Culturali](#) It offers news and information about exhibitions, events, archives, libraries and museums
- [National Museum of Science and Technology "Leonardo da Vinci"](#)
- [Nuovo rinascimento](#)
- [Photosearch](#) a collection of pictures
- [Sette lezioni online](#) History and literature of Southern Italy
- [Storia dell'Arte](#) personal site, created by a Liceo artistico teacher
- [The Italian Renaissance](#) Bibliography of academic sources on the Renaissance
- [Torino film festival](#)
- [Tourism in Tuscany](#)

##### Magazines

- [Art Online](#) Contains an updated section on art news, and a list of the main exhibitions. There are sections about old and contemporary artists

The Language Centre is happy to have increased the numbers of languages on offer. After ten year, it has gone from 80 to 200 languages and thanks to the great professional help of the Bodleian Libraries (in particular Bernadette O'Reilly, Bibliographic Standards Librarian and Chris Hargreaves, System Support Librarian), the collection is now catalogued and partly borrowable.

In truth, not all 200 languages on offer will have a lot of resources, and for some endangered or indigenous languages, we only have 1 or 2 items. This is why, perhaps, an online presence for all languages in the library is important. Indeed, we will not have much for Inuktitut as physical object (so far one DVD in the library) but we will offer several links online.

Inuktitut
^

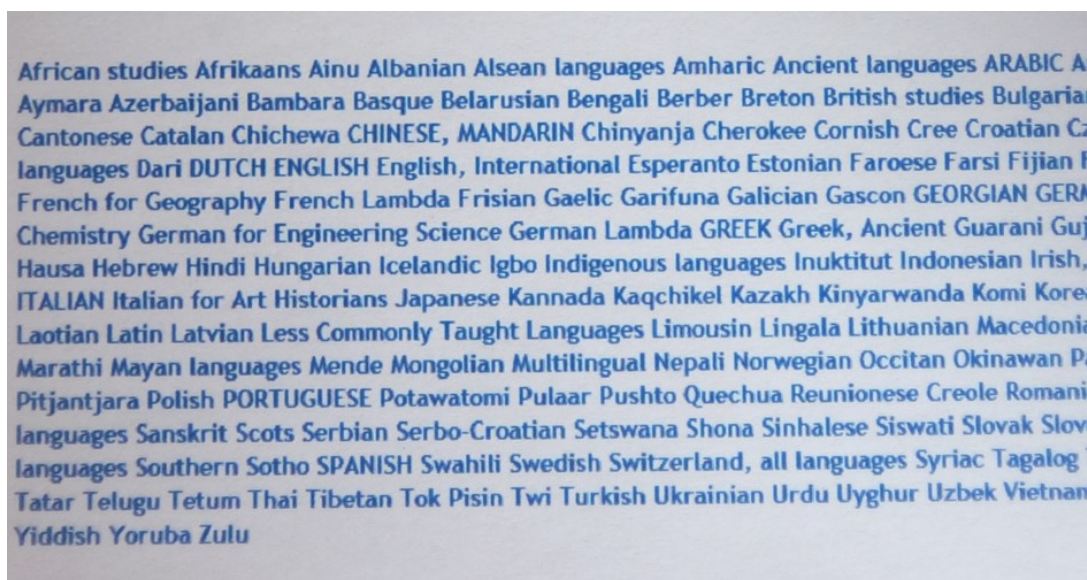
Library shelfmark: INU (only one film available)

- [Omniglot page](#)
- [Tussalanga.ca](#) learning (from English or French)
- [Livingdictionary](#)
- [CBC North](#) some radio programmes available
- [NAC Nunavut Artic College](#)

Likewise, many of the African languages books we have in the library date from missionary times. Yet, they can be the only printed sources available. Languages have changed a great deal since the publication of those books, and having online resources helps redress the balance, especially in terms of contemporary vocabulary and pronunciation.

In the summer of 2017, we were invited to participate in the creation of a new website, in partnership with [Oxford Mosaic](#), an Oxford University web making platform. Indeed, our previous website had had its days. And beyond the look, some facilities had gone from bad to worse such as the search box. One had to use google in order to find anything related to our languages' pages.

However, the languages' pages, had been classified in a very simple (if messy looking) way:



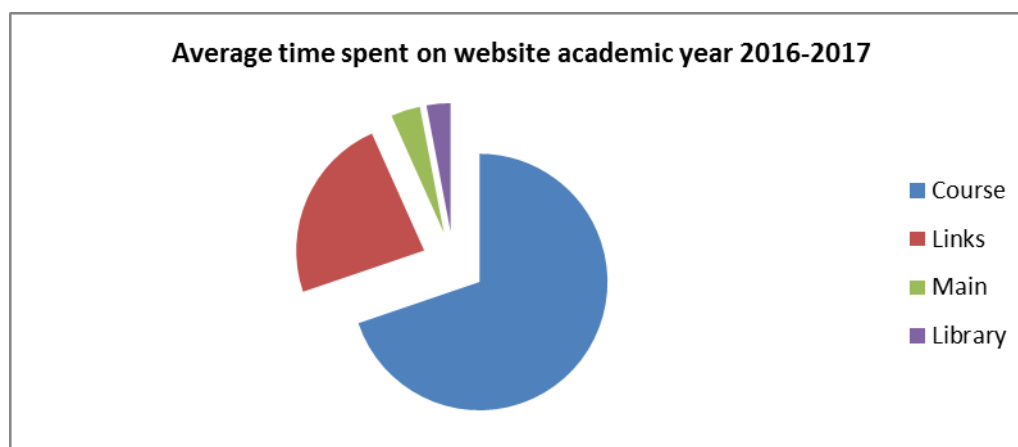
The Oxford University language centre website, details of the languages' page, 2010

The page was alphabetical by language, reflecting the library collection's classification, an in-house system, based on ISO codes, similar to the one used in [Cambridge University Language Centre](#) and in other substantial language centres such as [Glasgow University Language Resource Library](#). The classification on the website was alphabetical by the name of the language in English: GE for German for example, not Deutsch. You would click on the language which was hyperlinked, and you were taken directly to the language.

My role, in the making of the new website was to be a content editor only. I thought, at the time that the language learning links had had their days and that I ought to remove them altogether. I backed my decision on usage statistics data:

```
/links/japanese.html  
/links/urdu.html  
/courses/italian-closed.html  
/tests/tst_reqs_italian.html  
/links/irish_gaelic.html  
/links/portuguese.html  
/courses/english-pre-ses2.html  
/links/chinese.html  
/reg/i2.php?sm3=1
```

I transferred this data into an excel document, selected the first 100 and classified each page in the following categories: main information, courses, library, language links. And I was pleased to see that the language links pages were still being consulted. Though the views for the academic year 2016-2017 (October to June) were only 8% of the first 100 links for all categories, the average time spent was 23%.



Data from Google analytics on the first 100 pages of the website academic year 2016-2017

So the language links would stay. And in the middle of summer I was being offered, like a key to a new home, my pages already classified: but partly alphabetical, and partly geographical.



## Online Classifications

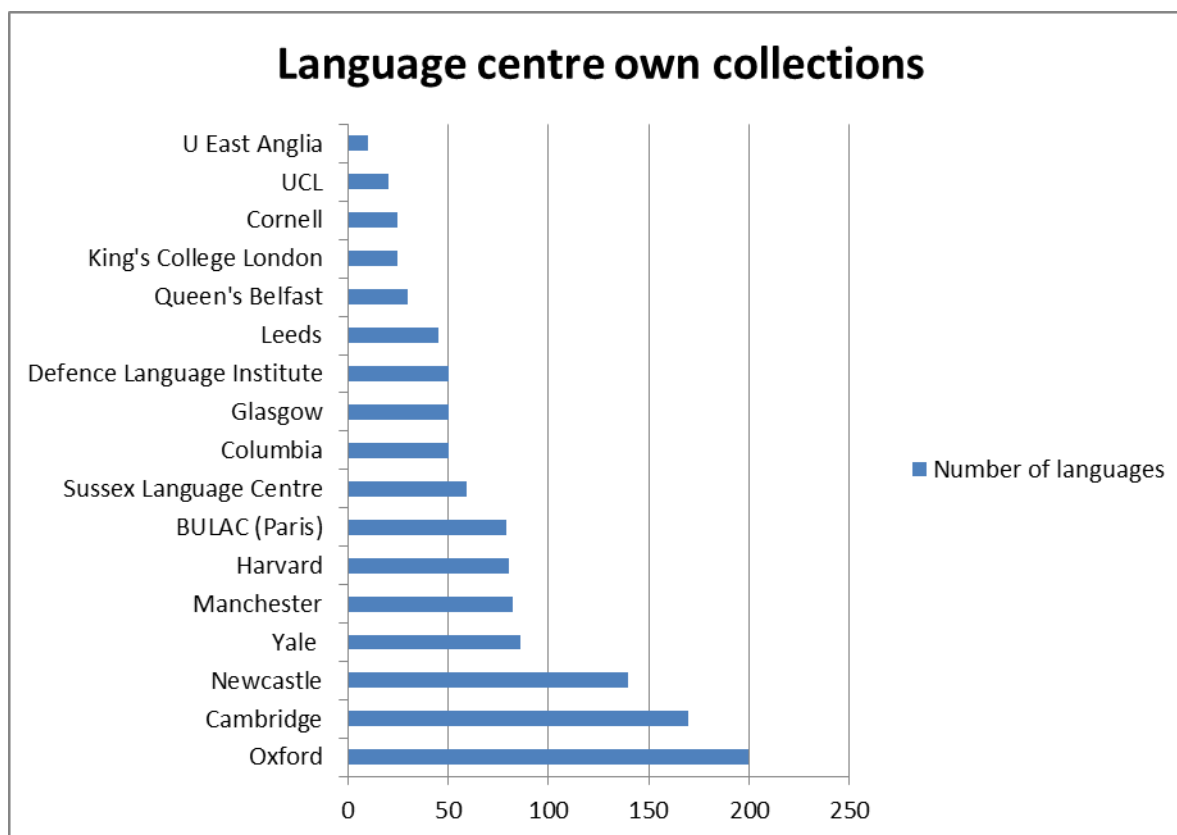
Languages can be classified in many ways. [Omniglot](#), a referential website when it comes to all things connected to language learning, presents several classifications.

A classification by [language families](#) could have been also possible, it would make sense to group the Turkic languages together in the website. Having said that, some families are enormous, for example the Indo-European family (455 languages or so), some families are small, even tiny, for example the Eskimo-Aleut family (11 languages). You also have the problems of many orphans... the language isolates, languages not linguistically connected with any others, Ainu and Basque for example. Like Simon Ager, the creator of Omniglot, you could create a group for all isolates, and another group for all creoles.

Another classification could be to [classify languages by writing systems](#). Then again, some groups (the Arabic script with Cyrillic with 115 language, the Latin script 711 languages) will be enormous while others just one unique system to write one language (Lao script for example). Le Monde Diplomatique shows us an interesting world map for writing systems [in this article](#)

## Other Language Centres

Having a look on the world wide web at other substantial language learning libraries I could see if other classifications were being used. I searched through members of [AULC \(Association of University Language Centres\)](#) as well as a few centres abroad.



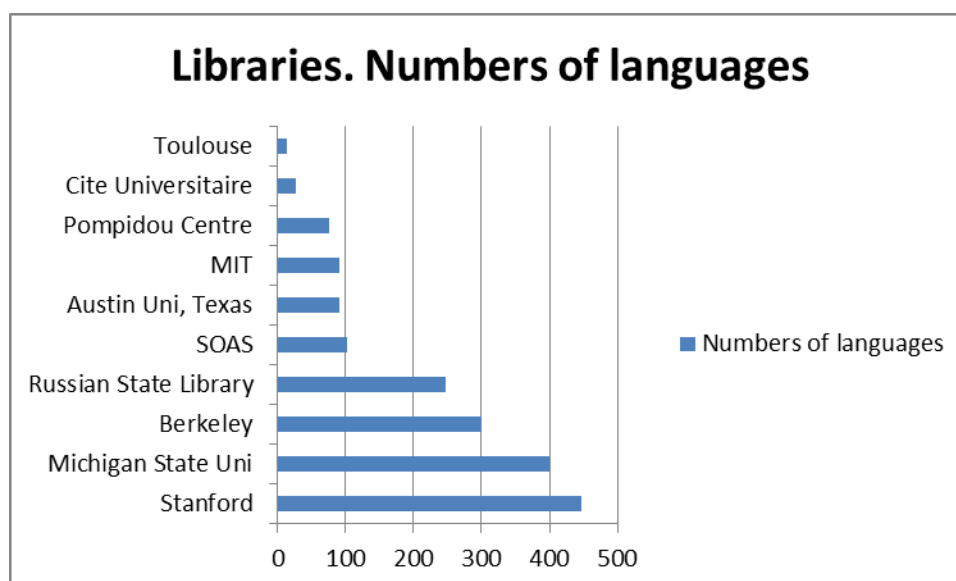
Data collected in 2017 (emails sent to several language centres). This is not, however, a ranking of language centre by language numbers but just a selection of language centres.

It was quite a surprise then to see that many links to external sites had gone from websites. Out of the 15 language centres shown above, 9 did not display any links anymore, 2 had one multilingual page only, and only 4 still had language learning links classified for 3 of them alphabetically. The fourth one geographically ([BULAC](#)).

Is the disappearance of external links pages because of the rise of VLEs? The online resources links chosen by language tutors will be specific to their students and not to all students and the general public. Or perhaps some centres wish only to show their own work, in particular open courseware ([Cambridge University Language Centre's open courseware](#) for example).

### Libraries with Substantial Foreign Language Collections

I then had a look at libraries with substantial language collections, though there is fuzziness in my finding. The materials are not necessarily for language learning, i.e. some libraries might simply have books in other languages not aimed at language learners.



Data collected in 2017 (emails sent to several libraries and catalogues' checks).  
This is not, however, an accurate view of the collections.

Same situation again: out of the 10 libraries in the graph, 8 of them did not appear to have language links. The two others were exceptions: [SOAS](#) with a geographical classification, then offering a selection of language links (similar to BULAC mentioned above) and [Austin University](#) with an alphabetical classification and quite fantastic open source collections.

### Rethinking the Classification

So I had to rethink the classification of languages from alphabetical to partly alphabetical and geographical. The already made pages were:

- All the languages we teach at the centre in alphabetical order (Arabic, Chinese, Dutch, French, Georgian, German, Greek, Italian, Japanese, Portuguese, Russian, Spanish)
- With the addition, in the same alphabetical classification, of other languages that had been considered important by the webmaster: they were languages that had, in the previous website, whole individual webpages (contrary to Less Commonly Taught Languages that were presented in one long html page)

Arabic	Bulgarian	Chinese (Mandarin)
Czech	Dutch	French
Gaelic	Georgian	German
Greek	Hebrew	Hindi
Hungarian	Irish Gaelic	Italian
Japanese	Norwegian	Polish
Portuguese	Russian	Spanish
Thai	Turkish	Urdu
Welsh		

Then I was offered the following:

Other African Languages	Other American Languages	Other Asian Languages
Other Australasian and Pacific Languages	Other European Languages	Other Middle Eastern Languages
Ancient Languages		
Ancient Greek	Latin	Other Ancient Languages
General Resources		
Indigenous and Endangered Languages	Multilanguage Resources	

Part of the new Language Centre website, November 2017

This is when the intellectual puzzles started. There were a few issues worth raising in order to offer the best classification possible:

- Is it the history of the language that matters? Is it where the language comes from, or is it where the language is spoken today? Do you place Pashto in "Other Middle Eastern Languages" as it belongs to the Eastern Iranian family, or do you place it in "Other Asian Languages" as it is spoken in Afghanistan and Pakistan?

- And speaking about actual geography, what are, exactly, Europe and Asia? There is quite a substantial amount of languages in between, such as in the [Caucasus region](#). Do I place Azerbaijani in Europe? I decided finally to place this language in “Other Middle Eastern” languages.
- Another geographical puzzle is to place languages that are not in the proposed classification boxes, such as island languages like Malagasi, from Madagascar, that I finally decided to place in... Africa, its closest continent.
- Where to place Esperanto, a constructed language? Europe, where it was created as its author, Zamenhof, is Polish? But surely, being a constructed language, it ought to have its own place. I had once a query from our Facebook page, about Enochian, the language of Angels and an email enquiring about one of Tolkien’s languages. Would we also place middle-earth languages in Europe then, as its author was living in Oxford? Clearly this would not make any sense. I could forget it altogether, yet given the queries I received, I felt I had to create a page.
- What about languages spoken by people who were displaced? [It is estimated by wikipedia](#) that the population transferred in the Soviet Union amounted to 6 millions. Amongst many others... would you place Crimean Tatar in Europe or in Asia?
- What about creoles and pidgins? Where to classify them? In the country where they are spoken now? In the country where they partly come from? For example put all English, French, Portuguese based creoles in the English, French and Portuguese pages? So Cajun (based on French) could be in Other American Languages, or in the French page?
- And the sensitive issues (cultural, historical)... Should I place the Armenian language in the “Other Middle Eastern” languages page? Or perhaps create a new “Eurasia” classification? I finally decided to place this language in “Other European Languages” My decision was partly taken after looking at the history of Armenia, the first country in the world to adopt Christianity.
- Due to several queries and some usage statistics regarding the British Sign Language, and the fact that we are starting to have substantial collections in the library, I thought it would be great to create separate pages for all the sign languages we have.

So I asked for the creation of the following three subsections:

Constructed Languages

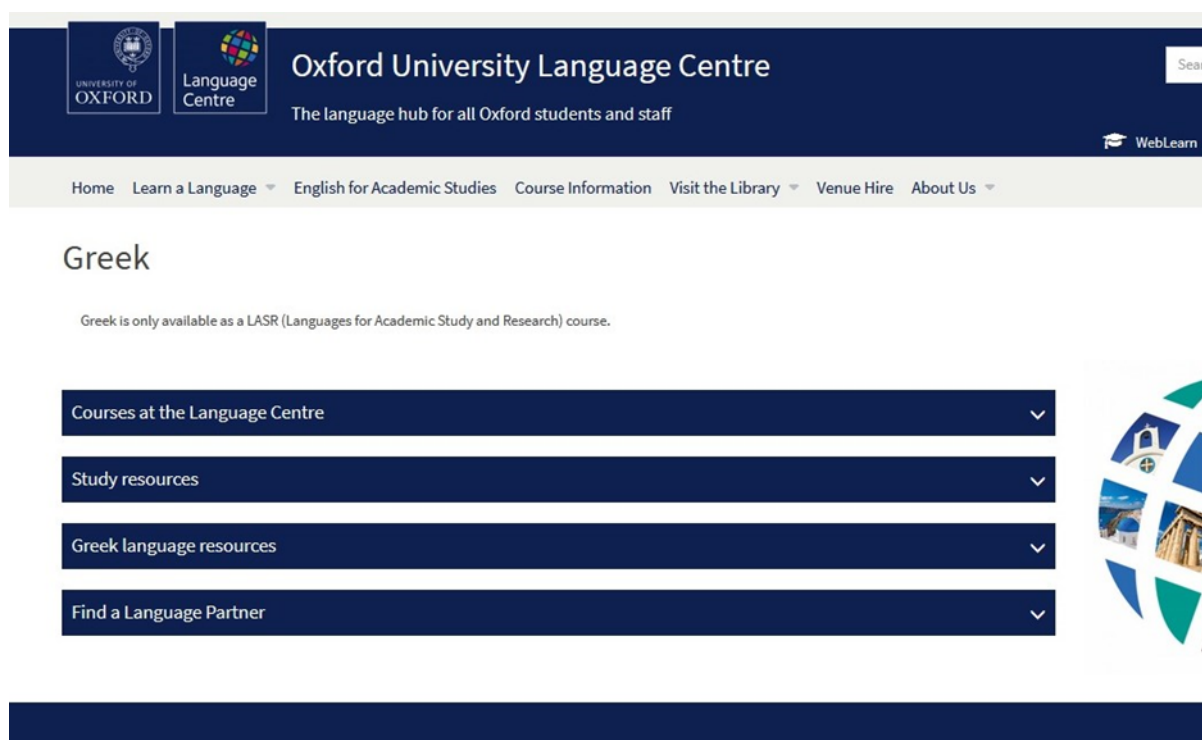
Creoles & Pidgins

Sign Languages

## Latest Statistics on Web Usage

The latest statistics from our website are quite different, given its structure and cannot be compared with the previous statistics. It is not possible so far to get statistics within a page. All languages taught at the Centre will have, on the same page, the language external links.

This is our Greek page for example:



Partial screenshot of the Language Centre Greek page, November 2017

So there is some fuzziness in data as both courses and links are in the same webpage. I have added a “course or link” parameter. Annual statistics should help define a clearer image and proper comparison could be done from the second year running of the new website.

## Conclusion

It will be the web user, of course, who will have the last word. So far there has been no criticism about languages perceived as misclassified. Should this happen, I would engage with readers and improve the situation, by either, reclassifying the language in question, reclassifying all non-taught languages or, in the case of getting statistics getting lower and lower, progressively remove all language learning links from our website, or move a selection of them to our VLE.

*Many thanks to Dr Jane Sherwood, Director of the Language Centre, and Bernadette O'Reilly, Bibliographic Standards Librarian, Bodleian Libraries, for their help and comments.*



Indexing languages constitute formalized languages designed and used to describe the subject content of documents for information retrieval purposes. In addition many scientific databases include a less controlled means of describing the subject content of documents: author-supplied key words. Gil-Leiva and Alonso-Arroyo (2007) conducted a detailed study of 640 scientific articles that possess author keywords and are indexed in databases. They found that author-supplied keywords have an important presence in the database descriptors studied: nearly 25% of all keywords appeared in exactly the same form as descriptors, while another 21% have undergone a normalization process but are still detected in the descriptors. In total, about 46% of the author keywords appeared in the same or a normalized form as descriptors leading Gil-Leiva and Alonso-Arroyo (2007, p. 1181) to posit that author keywords are a valuable source of information for indexing articles and for information retrieval.

In the last half-century, English has emerged as the dominant language of scholarly communication despite the fact that only about 6% of the world's population are Anglophones (Corcoran, 2015). What does this mean for scholars who are not native English speakers and may have only Limited English Proficiency (LEP)? Many study to achieve a high level of proficiency, while others may engage professional translation or editing services. Such options may be viable for well-funded scholars but may pose a challenge to LEP scholars from developing countries. In such cases, these LEP scholars may turn to cheaper alternatives, such as machine translation (MT), to help them engage in the scholarly communication process. In its 2013 *Trend Report*, the International Federation of Library Associations and Institutions (IFLA) lists MT as one of five key high-level trends in the global information environment. We are currently investigating the role of MT in various aspects of the scholarly communication process, including its potential for helping LEP scholars to understand scientific articles written in English, as well as to disseminate the results of their own research in English. However, our goal in this paper is more modest; we undertake a small pilot study to examine whether MT can help LEP scholars at the stage of searching. Kit and Wong (2008) observe that "MT quality is far from publishable," but they go on to suggest that MT may be good enough to serve many translation demands for the purposes of information access. While they did not test this application of MT, we will do so here in a small pilot experiment.

If we assume that many LEP scholars first learn about their domain through their own language and begin by accessing scholarly articles in that language, how can they take the next step of looking for comparable or related material in English? Can free online MT systems (e.g. Google Translate) help with searching?

### Corpus and Methodology

Some non-English-language journals provide abstracts and keywords in English. However, others, such as national journals or journals run by individual university departments – often by and for their own graduate students who are therefore new to the domain – may not. For this pilot study, we identified two Information Science journals in which the articles, abstracts and keywords, are provided only in Spanish. The first, entitled *Ciencias de la Información*, is a biannual online academic journal launched in 2011 and published by the School of Library Studies and Information Sciences at University of Costa Rica. It aims to promote research in a wide range of LIS-related areas including information systems, library studies, information policy, research methodology, and archival studies. The second is entitled *Métodos de Información*. Online since 2010, this biannual journal of the School of Library and Documentation Studies of Valencia addresses subjects related to libraries, archives and documentation centres.

For each journal, we randomly selected one article from each issue published in the past five years (2013-2017) for a total of twenty articles. From each article, we extracted the list of the author-supplied keywords, which were copied into a spreadsheet and sorted alphabetically. After eliminating duplicates, we were left with a list of 71 Spanish keywords which we then translated into English using Google Translate. While other free online MT systems are available, we selected Google Translate because Buitrago-Cirio (2018), who surveyed 48 LEP scholars at a university in Colombia, found that of the 42 who actively use MT, over 97% use Google Translate.

Next, we used the translated keywords to search the Library, Information Science & Technology Abstracts (LISTA) database. Although there are other bibliographic databases available for the LIS-domain, Vinson and Welsh (2014) report that LISTA has one of the broadest ranges, covering a wide variety of subjects pertaining to library and information science such as librarianship, bibliometrics, cataloging and classification, reference, online information retrieval, and information management. As Vinson and Welsh (2014, 124-125) emphasize, resources are a crucial consideration for libraries, and not every library can afford multiple databases. At institutions with limited means, which may be the case in developing countries, LISTA may well be the database of choice for its breadth of coverage and access to a variety of full-text materials. In addition, LISTA, without full-text availability, is offered free by EBSCO ([www.libraryresearch.com](http://www.libraryresearch.com)). We therefore elected to use the LISTA database for this pilot study.

Using the advanced search option, we restricted our searches in the following ways:

- Publication type: Academic Journals
- Document type: Article
- Language: English
- Field: SU Subject Terms (Performs a keyword search of subject headings, companies, people, and author-supplied keywords for terms describing a document's contents).

## Results

Of the 71 translated keywords, 37 (52%) returned relevant search results (i.e., articles on a similar topic to the corresponding Spanish article from which the original keywords were taken), while 34 (48%) did not. These 37 productive keywords appear to have been well translated.<sup>1</sup> Among the 34 translated keywords that did not return any results, 23 (68%) appear to be appropriately translated but simply not in alignment with the descriptors used in LISTA. See Table 1 for examples.

Keywords that returned results	Keywords that did not return results
<ul style="list-style-type: none"> <li>• <i>acceso a la información</i>: access to information</li> <li>• <i>alfabetización informacional</i>: information literacy</li> <li>• <i>Internet de las Cosas</i>: Internet of Things</li> <li>• <i>minería de datos</i>: data mining</li> <li>• <i>recuperación de la información</i>: information retrieval</li> </ul>	<ul style="list-style-type: none"> <li>• <i>comunicación científica</i>: scientific communication</li> <li>• <i>historia de la lectura</i>: history of reading</li> <li>• <i>libro impreso</i>: printed book</li> <li>• <i>patrimonio documental</i>: documentary heritage</li> <li>• <i>políticas de conservación</i>: conservation policies</li> </ul>

Table 1. Examples of translated keywords that did and did not return results.

1. This assessment was made by the present author, who is a certified translator.

For the remaining 11 (32%) keywords that did not return results, it appears that translation-related problems stemming from orthographic variation, synonymy, or differing syntactic preferences and semantic field coverage have interfered with the information retrieval process.

## Discussion

From our list of translated keywords, we can see that ‘ebook’ (*libro electrónico*) and ‘bibliographic data bases’ (*bases de datos bibliográficas*) use different orthographic variants than the LISTA descriptors, which are the full-form term ‘electronic book’ and ‘bibliographic databases’ (where ‘database’ is written as a single word). If the knowledge organization system (KOS) were more robust and able to handle such variants, then these translated keywords would have generated relevant results also.

Synonymy exists when two or more terms refer to the same concept. There were three cases where Google Translate translated a Spanish term into English using a synonym for the descriptor, rather than the descriptor term. For instance, *competencias informacionales* was translated as ‘information competences’ rather than as its synonym ‘information skills’ (which corresponds to a LISTA descriptor). Again, if the KOS could be expanded to better handle potential synonymic translations, then machine translated keywords could potentially generate more positive results.

Four of the translation problems result from the different syntactic structures most commonly used by Spanish and English. In all four cases, Google Translate has produced a literal translation that mirrors the underlying Spanish preference for prepositional phrases. The resulting translations are all grammatically and semantically correct; however, the more idiomatic way of expressing these structures in English is to use pre-modification. Moreover, in all four cases, if pre-modification had been used, the resulting term would have returned results from LISTA, as illustrated in Table 2. If a KOS could be augmented to recognize these very common and often predictable types of MT “errors” that arise from differing syntactic preferences between languages, then machine translated keywords could be more productive for information retrieval.

Original Spanish keyword	English translation by Google	Preferred English structure contained in LISTA descriptor
<i>arquitectura de la biblioteca</i>	architecture of the library	library architecture
<i>representación del conocimiento</i>	representation of knowledge	knowledge representation
<i>sociedad de la información</i>	society of information	information society
<i>utilización del espacio de bibliotecas</i>	use of library space	library space utilization

Table 2. Examples of differing syntactic structures

Finally, it is well known that languages divide the world up differently such that the semantic space referred to by a single term in one language (L1) might be covered by two different terms in another language (L2). In such a case translating from L2 to L1 is simple, but translating from L1 to L2 requires making a choice. For instance, the Spanish term *revista* can be translated as either ‘journal’ or ‘magazine’, and Google Translate chose ‘magazine’, which is the wrong choice in this particular context.

Similarly, the Spanish term *deontología* can be translated as ‘deontology’ or as ‘ethics’, and Google Translate chose to translate the keyword *deontología profesional* as ‘professional deontology’ rather than as ‘professional ethics’, which corresponds to a descriptor in LISTA. Because these choices are so often context dependent, it would be challenging for a KOS to deal with this issue. However, in our experiment, fewer than 20% of the translation issues fell into this category.

## Concluding remarks

While keeping in mind that this pilot experiment was conducted on a very small scale—using just two journals, 71 keywords, one MT system, one language pair, and one bibliographic database—the results nonetheless appear promising. Globally, only 11 out of 71 (15%) of the author-supplied keywords were translated in a way that led to no results being retrieved from the LISTA database. A much higher proportion of the keywords did not generate results because they did not align with the descriptors used in LISTA, rather than because the keywords were poorly translated. For LEP scholars, MT appears to be a viable tool for helping with information retrieval, and if KOSs could be augmented to better handle “predictable” translation-related challenges such as orthographic variation, synonymy and some types of syntactic variation, then MT could prove to be even more useful in this context. However, more work is needed to see whether MT can help LEP scholars to access and understand the content of the retrieved documents.

## References

- Buitrago Ciro, Jairo (2018). La littérature des sciences de la santé et la traductique: Quel est le rôle des bibliothécaires? Presented at the 2018 Canadian Health Libraries Association Annual Conference, St. John's, Canada, 15-18 June 2018.
- Corcoran, J. 2015. *English as the International Language of Science: A Case Study of Mexican Scientists' Writing for Publication*. PhD dissertation, University of Toronto, Canada.
- Gil-Leiva, I. & Alonso-Arroyo, A. (2007). Keywords Given by Authors of Scientific Articles in Database Descriptors. *Journal of the American Society for Information Science & Technology*, 58(8), 1175–87.
- International Federation of Library Associations and Institutions (IFLA). (2013). *Riding the Waves or Caught in the Tide? Navigating the Evolving Information Environment. Insights from the IFLA Trend Report*. The Hague: IFLA.
- Kit, C. & Wong, T. M. (2008). Comparative Evaluation of Online Machine Translation Systems with Legal Texts. *Law Library Journal*, 100(2), 299–321.
- Vinson, T. C. & Welsh, T. S. (2014). A Comparison of Three Library and Information Science Databases. *Journal of Electronic Resources Librarianship*, 26(2), 114–126.

Journals used as a source for the corpus

*e-Ciencias de la Información*: <https://revistas.ucr.ac.cr/index.php/eciencias>

*Métodos de Información*: <http://www.metodosdeinformacion.es/mei/index.php/mei>

### 1. Introduction

This report examines a small retrospective cataloguing project I undertook during a two-week student placement at Chatham House Library, in April-May 2018. The objectives of this case study are as follows:

- To provide a reflective context to develop some of my librarianship skills;
- To make suggestions for good practice in similar retrospective cataloguing projects;
- To consider possible implications for professional approaches to retrospective cataloguing and hidden collections

The next section outlines the background to hidden collections by drawing on some studies published in the last twenty years, and Section 3 describes in detail the project I undertook. Section 4 analyses that project with a focus on three challenging aspects, namely the initial appraisal of the hidden collection, the choice of appropriate cataloguing standards and the issue of institutional memory in a small organisation. Section 5 presents the report's conclusions and provides some recommendations.

As the original project was carried out in the context of a taught postgraduate programme, I had a dual role as learner and worker, and my work on the project should be contextualised accordingly. In other words, examples of 'bad practice' could be interpreted positively insofar as that practice has a pedagogical value. For this report's conclusions and recommendations, however, it is assumed that project workers would prioritise the organisation's objectives over their personal development. A draft of this report was made available to Chatham House library before submission.

### 2. Background

The issue of 'hidden collections' has been discussed widely within the librarianship profession in the last two decades. A survey of members of the Association of Research Libraries (ARL) in the United States in 1998 found that *'for printed volumes, about 15% of collections on average remained unprocessed or uncatalogued'* (Panitch, 2001: 49). An OCLC-led study of special collections and archives in the United States and Canada found (Dooley & Luce, 2010: 47) the same figure of 15% of printed books which were not included in online catalogues, and considerably higher non-inclusion rates for resources in other formats: 44% of archival material and manuscripts, 58% of cartographic materials and 71% of born-digital materials.

While the issue of hidden collections affects various kinds of libraries, it has often been studied in the specific context of archives and special collections, as was the case for the OCLC-led study mentioned above. A very similar study, with the same methodology and survey instrument, and indeed involvement from some of the same OCLC researchers, was then carried out in special collections and archives in the UK and Ireland: this time, the figure for printed books without online records was 22% (Dooley [et al.], 2013: 58). Just like in the American study, non-inclusion rates were higher for other formats, and the authors pointed to some important factors in relation to expertise: for example, they noted how *'for cataloguing of cartographic material, [...] expertise is rarely in-house in any but the largest institutions, and hence an area where further collaborative action would be beneficial'* (Dooley [et al.], 2013: 59).

The UK and Ireland study was able to draw not only on the OCLC methodology but also, significantly, on the findings of an influential study conducted only a few years earlier by *Research Libraries UK* in partnership with the *London Library*.



With data from 75 institutions, 60% of whom had retrospective cataloguing projects under way, that study had found a total of 18.5% uncatalogued volumes in those collections [4] (p. 6). Furthermore, they had noticed *'the presence of 21st century materials in the backlogs suggest[ing] that some libraries are unable to keep up even with current acquisitions.'* (Mertens [et al.], 2012: 40)

It seems then that retrospective cataloguing is indeed so common as to be considered an almost inevitable part of collection management. Many of those retrospective cataloguing projects are instances of retrospective conversion, i.e. the process of creating new versions of existing records in a different format (typically from card catalogues to computer records) and therefore can be seen as a by-product of libraries embracing electronic records in the last three decades of the 20th century. That historical reality, in turn, can create difficulties because potential funders may be unaware of that long lead time. As Dunia García-Ontiveros, who was in charge of retrospective cataloging at the *London Library* at the time of the study, pointed out, *'fundraising [was] particularly difficult as most donors thought that retro work was old hat. In general, donors were more interested in funding digitisation projects.'* (García-Ontiveros, 2010: 21)

All the studies mentioned above discuss the issue of the cataloguing standards adopted in retrospective cataloguing. The *RLUK-London Library* survey, for example, noted how *'historic practices in retro-conversion [...] leave a legacy of "minimal or inaccurate" catalogue records. These may be visible to potential users but may be misleading.'* (Mertens [et al], 2012: 23) This important aspect of cataloguing work will be considered later in this report.

### 3. The Project

The *Royal Institute of International Affairs*, commonly known as *Chatham House*, is a renowned and very active London-based think tank. It is a subscription organisation, with around 3,000 members and 160 staff, many of whom are active researchers. The library has a staff of three: a *Library and Information Services Manager*, a *Digital Resources Librarian* and a *Research Liaison Librarian*.

A particularly interesting feature of Chatham House Library is the way in which the library is at the core of the organisation, both because of its physical location within the building and the ways in which the library's space is used. It seems to be the natural choice for staff who want to escape their office and enjoy a quieter space to work, or have short, informal meetings with colleagues or guests. Library staff actively encourage that kind of use and, although there is clearly a cost to pay in terms of managing noise and expectations, the library's visibility within the organisation is truly remarkable. It is also clear that regular users have an excellent relationship with library staff and feel pride in having them as an asset.



Figure 1: Chatham House Library - Image rights: Chatham House (used by permission)

Chatham House Library has a clear emphasis on recent resources. The printed books collection on open access in the reading room dates almost entirely from the 21st century, whereas materials stored at other locations – while still mostly recent – include some items going back to the Institute's foundation in the 1920s. Some of the collection is stored offsite, at a location in Cheshire, and offsite items can be delivered on the next working day when requested before 4 pm.

The project I undertook concerned some of that material stored offsite. Two years ago, when the library first decided to use that location in Cheshire, some boxes (40 to 50, according to staff estimates) had to be moved quickly, and it became apparent that they contained items which had been overlooked during a previous retrospective cataloguing project. I catalogued the contents of 6 boxes, creating 40 new records and accounting for an estimated 15% of that hidden collection, which consisted almost exclusively of collections of treaties.

## **4. Some Important Challenges**

This section analyses some challenges encountered during the project in relation to three specific aspects. While no such treatment can be really exhaustive, those three aspects have been chosen to provide a basis for recommendations for good practice.

### **4.1 Initial appraisal of the hidden collection**

In an academic contribution focused on creating access to hidden collections, Barbara Jones suggested that libraries should '*develop different work flows for materials in various states of processing, for example, retrospective conversions records versus totally unprocessed materials.*' (Jones, 2004: 100) The retrospective cataloguing project at Chatham House did not fit exactly into either category: the items in the boxes had all been catalogued before, in the days of the old card catalogue, before an online catalogue was introduced in the 1990s. That old card catalogue was still used occasionally to identify and locate older resources, until 5 years ago, when it was disposed of. As the old records no longer existed, therefore, the resources clearly had to be catalogued anew.

As explained above, the items were stored offsite in boxes and, in the records I created, barcodes for individual boxes were used to record location. At the start of the project it soon became clear that, due to the sudden nature of the move to the new location two years ago, the boxes had been filled in an arbitrary fashion: different volumes of the same resource were sometimes found in different boxes. I therefore invested some time in inspecting the contents of all the boxes and rearranging some items before continuing with the cataloguing. In the specific circumstance of Chatham House, given the number of boxes and the cost of retrieving them from the location, that was probably the most efficient way of doing it. Other libraries, however, may be in a position to appraise an entire hidden collection systematically before the start of the project, depending on size, cost and location.

In one case, when cataloguing a multivolume resource, the fact that one volume could not be found in any of the 6 boxes prevented me from creating a complete record of acceptable quality. I therefore decided to invest some time in going to view the whole resource in the British Library's collection. Only later did I realise that a record for that resource already existed: a member of the library's staff had previously catalogued that 'missing volume', which was clearly in one of the other boxes at the offsite location. As a result, I had to amend the existing record.

I therefore started systematically checking for the existence of previous records before cataloguing each new resource. The insight proved useful for my attempts to develop my librarianship skills, by allowing me to reflect on the need to adopt a new general cataloguing rule, essentially requiring cataloguers to always check for existing records in advance.

In light of the findings of some of the studies cited above, for example on special collection in the UK and Ireland, which provide evidence of complex collections, often acquired, developed, managed and stored in idiosyncratic ways, that lesson is likely to be applicable to similar projects in many other libraries.

Finally, at the start of the project, library staff made it clear to me that I would be dealing with a low-priority and relatively low-value collection. At the end of the project, that assessment seemed accurate: none of the resources were rare and, even when made fully discoverable, they were unlikely to be of great interest to library users. Nevertheless, their marginal role within the library's collection could be used as an opportunity to highlight lesser known aspects of the library and improve its visibility. For example, although Chatham House Library claims that its collections date back to the 1920s, that hidden collection actually included some late 19th century items. Library staff capitalised on that unexpected find to further improve their visibility, as can be seen in the tweet below, with images of a pre-1920 resource. The novelty of that post may well have generated new interest even in seasoned library users.



## 4.2 Cataloguing standards

Chatham House uses its own classification system, as well as local cataloguing standards. In outlining the project to me, library staff made it clear that the resources in that hidden collection did not require very high cataloguing standards. Again, that seems to have been the right assessment as, given the fairly homogenous character of the hidden collection, and the very useful classification system, those resources are discoverable even with relatively minimal records. Nevertheless, the adoption of common cataloguing standards and the creation of richer records might create the conditions for potential collaborations with other institutions.

Tensions between short-term needs and long-term objectives, and between local choices and shared standards, are perhaps inevitable in projects of that kind.

The RLUK / London Library study cited above points out that *'the scale of the problem is often beyond individual institutions. Respondents support an online register of retrospective cataloguing and are interested in exploring national initiatives and technical solutions to bring this about.'* (Mertens 2012: 6) At the same time, several of those respondents argue that we should *'advocate, where appropriate minimal levels of cataloguing.'* (Mertens 2012: 42) Reconciling those two essentially contradictory approaches will require creativity and compromise. Ultimately, from an organisational management perspectives, decisions will have to be taken in relation to workflows and available resources (particularly human resources). In the case of Chatham House library, the decision to opt for local solutions seemed appropriate, yet it is difficult to evaluate the potential missed opportunities. For example, had I been able to draw on the kind of expertise about cartographic material mentioned in the study about special collections in the UK and Ireland quoted above, I might have found something of interest in the many maps included in those collections of treaties. In terms of specialist expertise, many libraries will be in the same position, and so this is arguably an area of librarianship where there is still scope for more cooperation.

One unexpected opportunity linked to the project at Chatham House was the availability of a new testing ground for the cataloguing system. The library had recently acquired a new, web-based, cloud-hosted Library Management System and was in the process of customising its interface. The cataloguing module allows for a very large number of fields and, as part of that customisation process, library staff had to decide which ones they would retain. Having the perspective of a new cataloguing project, with some specific features which were different from the usual cataloguing of new resources, provided extra input in that decisional process.

### 4.3 Institutional memory

The important contextual information I needed to work on those resources at Chatham House was gathered almost exclusively from the personal recollections of library staff. One of the three members of staff was particularly useful in that respect, as he was the only one to have witnessed important changes in the organisation since the 1990s, as in the case of the card catalogue being discontinued. His understanding of the importance of that hidden collection within the library's collections was unique. For example, he was aware of another, potentially more interesting hidden collection of newspaper cuttings from press digests which library staff used to produce. His assessment of the relative value of the collection I was working on was therefore irreplaceable; moreover, that assessment was based on information which was not available to anyone else.

Alison Cullingford, a leading specialist in special collections in the UK, lists five problems posed for libraries by hidden collections, including *'reliance on staff knowledge rather than catalogues: when staff leave, that information is lost.'* (Cullingford 2011: 84) The importance of cataloguing the stock therefore goes beyond the value or potential interest of a given collection: it can be part of a process of communicating institutional memory, including within the organisation. Catalogue records can essentially also be used as organisational records, and provide future organisational managers with the contextual information needed to take informed decisions. Adopting shared cataloguing standards could be a benefit, not just for potential partners and other institutions, but also for an individual library's own records management practices. Once again, organisational decisions about workflows and resources will be crucial; many relatively small organisations, such as Chatham House, will probably conclude that their resources are limited, and that kind of cooperation is therefore not achievable, so this seems likely to remain a significant obstacle in dealing with this aspect of librarianship.

## 5. Conclusions and Recommendations

The retrospective cataloguing project at Chatham House provided a useful reflective context for developing my librarianship skills. For example, I gained the simple but important insight that, as a cataloguer, I should always check that the next resource in my in-tray has not been catalogued already. It was also an opportunity to consider wider issues about managing workflows in dealing with hidden collections, particularly in relation to choosing cataloguing standards and identifying opportunities for collaborating with other libraries.



Although it is difficult to generalise based on one small project, the Chatham House experience showed an example of the limits of retrospective conversion: cataloguing the resources anew was the better option.

The following recommendations are offered to libraries which are considering retrospective cataloguing of hidden collections:

- Libraries should consider the opportunities provided even by small investments in retrospective cataloguing projects: for example, newly discovered aspects of the collections which can be used for outreach projects or to improve visibility.
- Libraries should also consider the opportunities for the organisation's internal processes, for example in reviewing their own cataloguing practice and in contributing to records management at organisational level.
- Investment in the initial appraisal of hidden collections is likely to make the process efficient and more easily manageable at later stages: if circumstances allow it, a systematic appraisal of the whole hidden collection before cataloguing starts is recommended.
- Decisions about cataloguing standards will have to reconcile local realities with the opportunities to share records and make resources discoverable even more widely: whilst mindful of constraints, libraries should generally aim for richer records.

Libraries should aim to foster a culture of collaboration encompassing other aspects of the process, for example by participating in the creation of registers of retrospective cataloguing, which is likely to help smaller organisations to maintain and communicate their institutional memory.

## References

- Panitch, J.M., 2001. *Special collections in ARL libraries: Results of the 1998 Survey*. Washington, D.C.: Association of Research Libraries. [online] Available from: <http://www.arl.org/storage/documents/publications/special-collections-arl-libraries.pdf> [Accessed 30 May 2018]
- Dooley, J. M., and Luce, K., 2010. *Taking our pulse: The OCLC Research survey of special collections and archives*. Dublin, Ohio: OCLC Research. [online] Available from: <http://www.oclc.org/research/publications/library/2010/2010-11.pdf> [Accessed 30 May 2018]
- Dooley, J. M., Beckett, R., Cullingford, A., Sambrook, K., Shepard, C., and Worrall, S., 2013. *Survey of Special Collections and Archives in the United Kingdom and Ireland*. Dublin, Ohio: OCLC Research. [online] Available from: <http://www.oclc.org/resources/research/publications/library/2013/2013-01.pdf> [Accessed 30 May 2018]
- Mertens, M., García-Ontiveros, D., Cullingford, A., and Jeskins, L., 2012. *Hidden collections: report of the findings of the RLUK Retrospective Cataloguing Survey in association with the London Library*. London: Research Libraries UK. [online] Available from: <http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Hidden-Collections.pdf> [Accessed 30 May 2018]
- García-Ontiveros, D., 2010. Retrospective cataloguing: The forgotten projects. *Catalogue & Index*, 161, 21-25. [online] Available from: <https://archive.cilip.org.uk/sites/default/files/documents/CandI161.pdf> [Accessed 30 May 2018]
- Jones, B. M., 2004. Hidden collections, scholarly barriers: Creating access to unprocessed special collections materials in America's research libraries. *Journal of Rare Books, Manuscripts and Cultural Heritage*, 5(2), 88-105. [online] Available from: <https://search.proquest.com/docview/57606165?accountid=14511> [Accessed 30 May 2018]
- Cullingford, A., 2011. *The special collections handbook*. London: Facet.



Thank-you so much for sending me to the 2018 CILIP Conference. I recommend attendance of CILIP Conferences – get it into your CPD plan now.

I found several threads of topics likely to be of interest to CIG members, so perhaps I can weave for you a conference tapestry. I may now leave that metaphor behind in case this account is too pervaded by warp and weft...from two very interesting and enjoyable and intense days, I am paraphrasing, altering the sequence and giving much-reduced extracts.

In fact, some of the important items I am going into very little depth at all about are: [Sally Walker](#) (award-winning librarian) on her many activities as Children's Librarian in Orkney; [EveryLibrary](#) who are very effective in applying an activist model to money-raising and support for libraries in the US; CILIP President Ayub Khan's breakfast session on an international strategy (so many aspects including an opportunity to get involved in the [Presidential Commission](#), perhaps by completing the [survey](#)); Samira Ahmed, journalist and broadcaster, on how the Windrush scandal looks to her (she feels genuine astonishment that the landing cards weren't given to the National Archives); Helen Dodd of Cancer Research UK on their journey to compliance with GDPR; TFPL on their survey of senior knowledge and information managers with amongst other things, a count of people saying they use AI for automating document review, content extraction and search activities; Dr Katharine Schopflin on working in a corporate situation, not necessarily a library, as an information professional; a session on the skills to be gained by volunteering chaired by Helen Berry. Alas, I can't tell you about all the parallel sessions that I could not attend!

They are cataloguing and indexing in the [House of Commons Library](#) as one part of a vital information ecosystem described for our benefit in a fascinating talk by Penny Young, House of Commons Librarian. Their first library catalogue was created in 1830; we were told that although relatively well-resourced, the library is none-the-less having to manage demand, particularly at the moment. MPs are busy, hard-working people with a difficult job (and I believe this was something she wished to convey in these anti-expert times, however tempting it might be to be cynical), supported by the library with confidential enquiries/research (30,000 p.a.) and briefings that are available to the public (900 p.a.). The library is a trusted brand.

Cataloguing and indexing (and abstracting and classification and many other tasks relating to information management) also occur in the information team of Linklaters, an international law firm: CJ Anderson, the manager of 96 information staff globally gave a talk on the evolution of their information service and the impact of AI upon it. Other professionals (e.g. lawyers, marketing) are beginning to ask the information team to the table not least on the basis, I think I understood, of the corporate taxonomy they created and that is used in their knowledge and information management (K&IM). At some point when she and her growing team had accumulated just about every task relating to information in the company (I am slightly awe-struck at that), they sat down and worked out what they wanted to keep, and what they did not (which incidentally I think must also have involved some more classifying, of tasks). Ms Anderson states that an information team don't need to be technologists: they just need to be able to see what the technology can do.

The new knowledge management accreditation was one of the things featured in Nick Poole's inspiring high-level overview on CILIP's activities for those of us who got here on the Tuesday evening before conference itself started (and thank-you, South East Branch, for your hospitality!). It is a good fit for CILIP, who already have an established accreditation process and is perceived as a complimentary accreditation that CILIP is well-capable of handling, which is needed (because at the moment anyone can call themselves a knowledge manager) and which strengthens CILIP.

In the interesting K&IM session on Wednesday Paul Corney included the curation of information as one of several critical skills required by what he calls a 'Kowledgeur', someone who takes knowledge gained from one activity or market, and uses it elsewhere. There's more about his Kowledgeur concept on [Knowledge et al.](#) There is an ISO standard on knowledge management soon to be published. Stephen Phillips of Morgan Stanley and the K&IM SIG in his clear presentation on the K&IM portfolio points out that now is a good time for the knowledge management initiative, which will be a quality mark to work towards, and also give recognition of knowledge management skills for the many people working in the field. There will be three levels (certified, chartered and fellowship).

Stephen also talked about the time being right to revisit the 1994 Hawley Report which covers the value of information to an organisation. The updated report is scheduled for publication in September 2018 after the opportunity to participate was mentioned at conference: [Information as an Asset](#). That report was also one of just four influential reports featured in Guy Daines' interesting presentation on the occasion of his retirement (a transcript has been provided [here](#)); the others were the 1993 [Follett Report](#), the [New Library: The People's Network](#) and Hargreaves' 2010 [Digital Opportunity](#). Along with several other speakers, he also mentioned ethics including the [ethics review](#) being worked on by CILIP.

In a breakout session with CLSIG, Oz Ablett from AECOM introduced us nicely to BIM (Building Information Management) with its 'combined view' of a structure with all relevant information, and the need to get the words right (for example because the different professions involved use different vocabularies) and wonderful outputs (e.g. virtual/augmented reality) – I am paraphrasing quite a lot here, as the presentation involved a lot of cheery images of Lego! In broad summary: vast amounts of information go into the planning, creation and lifetime of a building and BIM holds all of this together; BIM is mandated by government for public projects.

In the technical practice thread David Haynes of [ISKO](#) and City University gave a wide-ranging talk around issues of knowledge organisation, privacy and fake news, pointing out that indexing policy can affect access to information and there is no way of impartially conveying information (for instance, think DDC). Metadata is used in several activities of topical interest including the way people describe themselves (e.g. on social media), surveillance (e.g. [Greenwald, 2013](#)), the 'filter bubble' (he cited Pariser (2011) – I found the book, and there's a [TED Talk](#)) and digital forensics. He talked about the work of ISKO and I have also recorded mentions of Haynes' own work (eg [Haynes, 2018](#) and [privacy calculus](#)) and a [Post-Truth Forum](#).

In an informative call-to-action, Catherine Cooke of [BIC](#) talked about a set of interoperability standards replacing SIP2 for connections between the LMS and other library applications. She said [LCF](#) is designed to replace SIP2 and will for example enable connecting the LMS with computer booking and print accounts as well as self-issue machines, so that library users get information about their fines and PC bookings and printing charges at the same time when they access the system, and the administrative overhead is reduced. Catherine Cooke mentioned that there are many other functionalities that can be added, with lots of ideas for all public libraries – but librarians need to ask their suppliers for LCF. She points out that librarians don't need to understand the technicalities, because that is the job of the suppliers. She says they can ask suppliers if they are on the LCF development panel; if they aren't, or even worse haven't heard of LCF, Catherine says that tells you something about your supplier. She highlighted that BIC has a track record in developing standards for libraries and six public authorities have agreed to pilot LCF.

Dr Deborah Lee of the Courtauld Institute of Art presented a practically-minded discussion of cataloguing. She commented that cataloguing might now be known as resource description or metadata, although the people that do it tend to call themselves cataloguers (although also noting it's not always in the job title). She pointed out that it is all to facilitate retrieval and also provides an inventory of a collection. Dr Lee said cataloguing is done for several user groups in the library (patrons, colleagues in several roles) and even outside the library, and relationships are needed with all users. She stated that if the main user is local, local requirements need to be the primary focus. A description of the many daily tasks of a cataloguer and of what they think about as they catalogue was provided.

Local requirements, interpretation of RDA to fit local information and display requirements and ethical considerations all featured. She asked, for instance, if an artist depicts something that never happened, should the catalogue inform the user of that? Then in a splendid burst of CPD-inducing acronyms, Dr Lee mentioned some future developments ([3R project](#), [LRM](#), [BIBFRAME](#), linked data and increased ability to manipulate data). She also offered us some of her thoughts on the recruitment of capable cataloguing staff and a wish-list for the future.

Alison Brett (University of Salford and LIRG) provided us with a couple of activities which got us involved in a session on 'the evidence in evidence-based practice', and mentioned that CILIP is promoting evidence-based practice and is setting up a portal pulling together evidence for the use of librarians. We heard Dr Diane Rasmussen Pennington give a comprehensive overview of who, what, why and how to do practitioner research (with another acronym for me (EBLIP – evidencebased library and information practice)), mention some specialised techniques including transcription and content analysis, statistics, secondary analysis and data mining; and mention that LIRG have put an introduction to research methods on the VLE. [Library and Information Research](#) is their open-access journal. It seems relevant to note that her first piece of practitioner research was while working as a cataloguer. Jenny Turner from East Sussex Healthcare talked about what sounds like very beneficial work collecting and using evidence routinely across health libraries: a study was conducted that gave rise to a Toolkit to help such research. Leo Appleton kindly told us about his part-time PhD at Edinburgh Napier University, using a longitudinal focus group methodology to measure the value and impact of public libraries. He has been re-visiting a focus group over time to see how things change. He should complete his PhD fairly soon; the research questions changed through the design, and if using a focus group he recommends a pilot first. He mentioned that he codes his research results to be able to analyse them. Part of the conclusion of this session was that anyone can do some practitioner research. It's not just for geniuses! My pre-conference literature search had caused me to think that it would be of interest and relevance to CIG members for me to attend the session. I would like to think (although I know this is not evidence!) that the analytical thought habits developed in CIG-related procedures would also find good application in research; many meta-analyses in medicine following approved methods refer to the 'advice of an experienced...librarian' for search strategies, which are part of research and are the other side of the indexing 'coin'.

The links in this report were found, accessed and believed to be sensible and useful, on 20th August 2018.

The editors welcome letters on relevant topics and responses to articles in previous issues. Please email us with your comments.

### **Cataloguing and the Greatest Good / Arwen Caddy (C&I, issue 191 (June 2018))**

This article was a most interesting read, giving lots of food for thought. I was reminded of S.R. Ranganathan's Five laws of library science throughout, which, to my mind at least, are concerned with the greater good of the library, by satisfying the user in their search and retrieval. This linking with Ranganathan was particularly strong for me in section 5, Consistency and Perfection?, where Caddy discusses the cataloguing practices of locking down a record as soon as it is created. If a mistake is made, or a more appropriate access term becomes available, then my advice would be to use both terms in newer records, allowing the user to conduct their own cross-reference searching. If this doesn't satisfy the rule about the library being a growing organism, I don't know what does! Consistency can be static, and doesn't allow for growth and change; be adaptable (even the most consistent of cataloguers need room to grow). Caddy started with a Pixar quote; I'll finish with a slightly different film (which, at the heart of it, is also about ethics): Life [or metadata] finds a way.... (Jurassic Park, 1993).

Natasha Aburrow-Jones, SUNCAT Project Officer



**Catalogue & Index** is electronically published by the Cataloguing and Indexing Group of the Chartered Institute of Library and Information Professionals (CILIP) (Charity No. 313014)

**Advertising rates:** GBP 70.00 full-page; GBP 40.00 half-page. Prices quoted without VAT.

**Submissions:** In the first instance, please contact the Co-editors:

Karen Pierce: [PierceKF@Cardiff.ac.uk](mailto:PierceKF@Cardiff.ac.uk)

Deborah Lee: [Deborah.Lee@courtauld.ac.uk](mailto:Deborah.Lee@courtauld.ac.uk)

**Book reviews:** Please contact the editors

ISSN 2399-9667

CIG website: [cilip.org.uk/cig](http://cilip.org.uk/cig)

CIG blog: <http://catandindexgroup.wordpress.com/>

CIG twitter: @CILIPCIG

Tags from the CIG blog on specific areas of interest:

[authority control](#) [book reviews](#) [Catalogue and Index](#) [cataloguing](#) [CIG activities](#)  
[CIGS](#) [classification](#) [committees and working groups](#) [conferences](#)  
[Dewey](#) [digitised material](#) [Dublin Core](#) [events](#) [folksonomies](#) [linkblog](#) [MARC](#)  
[metadata](#) [news](#) [RDA](#) [Semantic Web](#) [social software](#)  
[standards](#) [taxonomies](#) [UDC](#)