FAST Forward? Sustaining and extending subject indexing at the British Library

Alan Danskin, Collection Metadata Standards Manager, British Library

FAST stands for Faceted Application of Subject Terms.[1] 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, Bodeleian 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.[2] The survey posed three scenarios for comment:
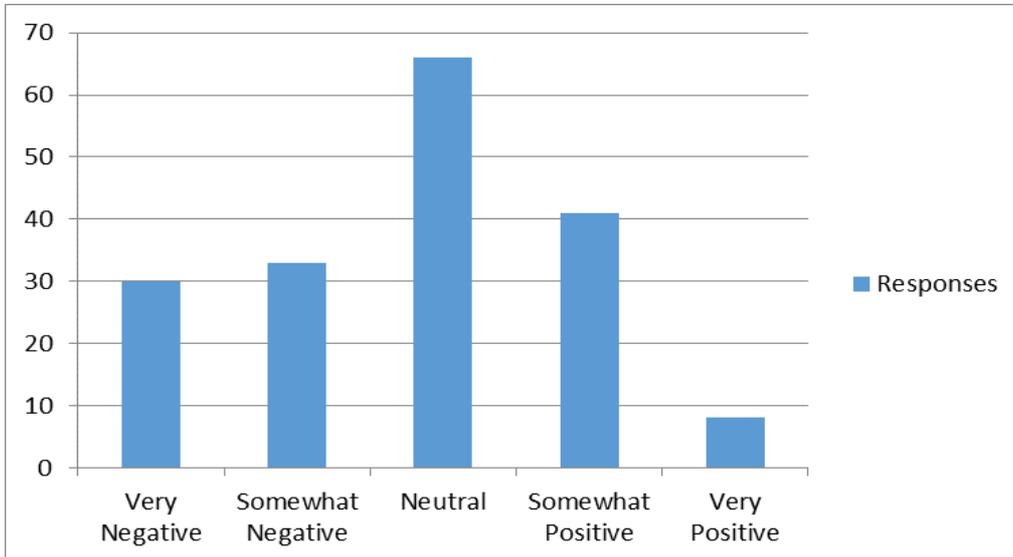
1.The British Library proposes to adopt FAST selectively to extend the scope of subject indexing of current and legacy content.

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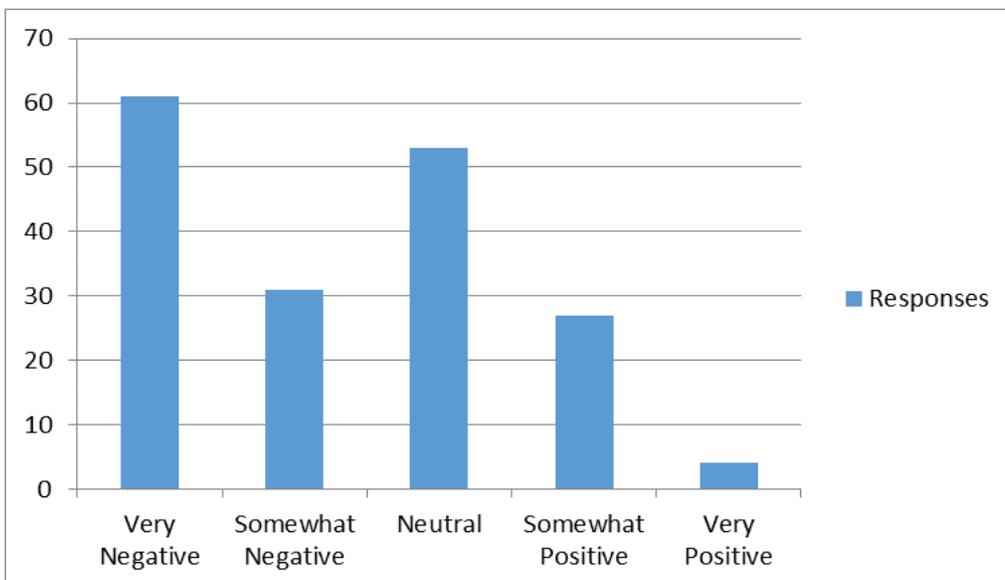
Proposal 2 The British Library proposes to adopt FAST selectively to extend the scope of subject indexing of current and legacy content

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.[3]

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.
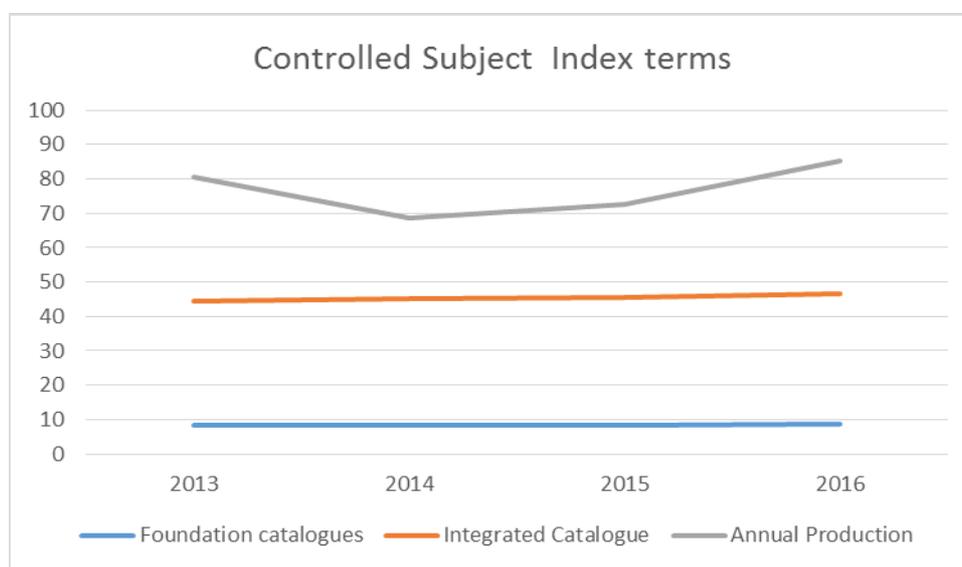

Figure 1 Controlled Subject Index Terms in catalogue production db

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 finding were published in February 2018.[4] 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.[5]

---

4. Survey on "Next steps for FAST": summary of results, conclusions and next steps. February 8, 2018. 8https://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/1XMiaj91qsTKFQIrhRaWHpxyHAPItR-ixlFpR7dbZtLw/edit#slide=id.p3

5.Janet Ashton & Caroline Kent (2017) New Approaches to Subject Indexing at the British Library, *Cataloging & 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.