Thema: the subject category scheme for a global book trade

Thema is the new global subject classification for books and e-books, which has already gathered wide international support.

Background

Subject classification of books and e-books is one of the keys to improved merchandising, discoverability, sales and market intelligence.

Prior to the development of _Thema_, subject categorization schemes relevant to the book trade have been nationally-focused. For example, the North American trade uses the BISAC scheme administered by the Book Industry Study Group (BISG). The UK trade uses a scheme devised by Book Industry Communication (BIC). The remainder of the English-language publishing world is forced to use both, with the added cost and complexity that implies. Whether it is CLIL in France or WGS in Germany, each country has its own indigenous subject classification scheme (and often more than one…). In contrast, established global schemes – like Dewey or the Universal Decimal Classification (UDC) – are used primarily in the library world and are unsuitable for the needs of the book trade.

Until 2010, these national divisions in the book trade were accepted as inevitable. Within any one country, the schemes were refined and optimized. Publishers and booksellers operating internationally implemented a complex 'cat's cradle' of mappings from one scheme to another. But each book classification that's mapped loses quality, and mappings require maintenance as schemes are revised. At the same time, an international metadata scheme like ONIX for Books allowed most other key metadata elements to be conveyed relatively smoothly between organizations, across national and linguistic barriers.

So the value of a _global_ classification scheme is clear: it lies in reduced duplication of work where more than one scheme is in use, and the elimination of costly and imprecise mapping processes – and this value grows with the increasingly international nature of the book trade.

Origins and aims of _Thema_

The origin of _Thema_ lies in an attempt to 'internationalise BIC'. The BIC subject category scheme – launched originally in 1997 – is clearly and deliberately 'British' in flavour, and tuned to the needs of the UK book trade. But modified versions of the UK BIC scheme were already in use in Italy, Spain and other countries. During 2011 and early 2012, a proposal for a more internationally-balanced version of the BIC scheme named IBIC emerged. And while IBIC itself was never released, it led to

the formation of a much larger group of stakeholders – including BISG and BookNet Canada from North America – willing to work on a global scheme that ultimately became *Thema*.

Thema aims to be global in scope, multi-cultural and multi-lingual, applicable to all parts of the book supply chain, and flexible enough to allow each market to retain its unique cultural voice, while remaining a unified and simple-to-adopt standard. It can be used alongside existing national schemes like BIC, BISAC, WGS or CLIL, and has the potential to eventually replace them – though this is not an immediate goal.

The *Thema* project was formally launched at a EDItEUR session during the Tools of Change Frankfurt conference in October 2012. It garnered support from trade organizations in more than a dozen countries, crucially including both North American and European stakeholders. A pilot version of the scheme was developed and published, and a governance structure was built to ensure a secure future for the scheme. And Frankfurt 2013 saw the formal launch of version 1.0 of *Thema* in an initial four languages, plus the first governance meeting of the *Thema* International Steering Committee.

Experience gained from translation of the scheme, from the creation of mappings from legacy schemes, and from the first year of real-world use by early adopters informed the creation of *Thema* version 1.1, which was agreed by the International Steering Committee at Frankfurt 2014 and published in November 2014. A second update – *Thema* 1.2 – was published in May 2016. Note that all *Thema* 1.0 codes are retained – later versions consist of additions only – so updating does not require much work and existing metadata does not need to be revised.

Thema is intended as a subject scheme for use within the metadata supply chain from publisher to retailer or to public library. It is not always expected to be *directly* visible to the consumer, but may be mapped by retailers into their own proprietary 'browse by subject', search terms or shelving schemes.

Governance and status of *Thema*

Thema is managed by EDItEUR, the international trade standards body for the book trade, which also manages ONIX for Books, EDItX and other metadata, identifier and e-commerce standards. EDItEUR is a not-for-profit membership-supported organization that works for the benefit of the global book trade. It has 100+ members in more than 25 countries around the world.

Following the successful ONIX model, EDItEUR has set up an International Steering Committee for *Thema*, comprising representatives of various national stakeholder groups, to guide the future development of the scheme. The committee will meet face-to-face at least twice a year, during the London and Frankfurt book fairs.

Like other EDItEUR standards, *Thema* is free of charge for anyone to implement, under a permissive EDItEUR license (DOI: 10.4400/nwgj). In particular, there is no requirement for EDItEUR membership – though membership is always welcomed, and members have a more direct stake in the future development of EDItEUR's standards.

Since it was launched, *Thema* has been adopted as the preferred subject category scheme by a growing group of organizations – individual publishers, bibliographic data aggregators, books-in-print registries and retailers.

What steps should a publisher, retailer or trade organization take?

- Publishers concerned with the difficulties of using multiple national subject classification schemes, or the cost of maintenance and imprecision of mappings from scheme to scheme should consider implementing *Thema* alongside their own national or internal subject scheme, and embedding the *Thema* codes in the metadata they distribute to their supply chain partners. The global nature of *Thema* makes this particularly important for publishers who expect growth in the e-book market or who trade with international e-book retail platforms, or who export a significant proportion of their physical products.

- Retailers selling books or e-books that are traded internationally should consider making use of *Thema* as a source of subject information, to drive 'browse by subject' and search functionality on websites and in internal systems, and to guide physical merchandising plans.

- Informal groups of *Thema* users, formal publisher or bookseller associations or other interested trade bodies should consider facilitating a *Thema* national group – a stakeholder group represented on the International Steering Committee. In order to ensure EDItEUR has the resources to manage the scheme properly, at least one member of the group should be an EDItEUR member, but EDItEUR places few other requirements on the nature of such a group. As well as being represented on the steering committee charged with guiding overall development of the scheme, national groups have an important role in creation of translations of the scheme, mappings from existing national subject schemes to *Thema*, and of *Thema* 'national extensions' – optional parts of the scheme tailored to local detail that may be unnecessary at the global level.

Technical structure of the *Thema* scheme

Although it draws most heavily on BIC, *Thema* builds upon the experience in many established national schemes. It lays out a hierarchy of subjects: 20 top level categories, each subdivided into many sub-categories, and each having a heading (*eg* 'Geological surface processes (geomorphology)'), an alphanumeric code (eg 'RBGD'), and in some cases associated notes. There are around 2500 subject headings in total, and some examples are shown below:

Code	Heading
AGA	History of art
XAMC	Manga: Kodomo
NHW	Military history
QRRF	Zoroastrianism
KJMP	Project management
LWKF	Shariah law: family relations
MKE	Dentistry
UGB	Web graphics & design
WBB	TV / celebrity chef cookbooks
YBC	Children's picture books

The hierarchy implies that books should be classified to the most detailed appropriate level (not necessarily the most detailed *possible* level), and a book classified as – for example – AGA (History of art) is also automatically and more generally associated with the AG (Fine arts: treatments & subjects) and A (The arts) categories.

In addition to the headings, there are post-coordinated 'qualifiers' that can be used to refine the meaning of the main subject categories, for example:

Code	Heading	Qualifier type
1HFGU	Uganda	Geographical
2ACSC	Icelandic	Language
3MD	16th century, c 1500 to c 1599	Historical period
4GH	For International GCSE (IGCSE)	Educational purpose
5AG	Interest age: from c 6 years	Interest
6BA	Baroque	Artistic style

A note with the AGA (History of art) category suggests it should be used in conjunction with Style, Geographical and Historical period qualifiers, to refine the classification as appropriate.

And finally there are 'national extensions' within the qualifiers. The extension codes are arranged so that national extensions can be truncated to leave a meaningful 'global level' qualifier:

Code	Heading	Truncated	Heading
1DNS-SE-BH	Västergötland, Sweden	1DNS	Sweden
3MPBGJ-ES-B	Spain: Civil war (1936–1939)	3MPBGJ	c 1930 to c 1939
4Z-UK-SD	For Scottish Curriculum National 5	4Z	For specific national curricula
5HC-US-A	US Independence Day	5HC	Holidays and celebrations

In principle, the headings and any associated notes can be expressed in any language – dentistry (MKE) can also be Zahnheilkunde, odontología or طب الأسنان – without the meaning of the category changing. *Thema* is available in around 20 languages so far, but the codes remain language-independent. Only the codes are communicated in common bibliographic metadata, which means the recipient can understand the subject of the book in their own language.

Code	Heading
CF	Linguistics
CF	Sprachwissenschaft, Linguistik
CF	Linguistique
CF	اللغويات
CF	言語学

And as well as the codes shown in the examples above, *Thema* categories and qualifiers will also have persistent URIs that can be used in linked data applications, and the scheme will be published online as SKOS.

One book may be classified in several categories if necessary, and with several qualifiers, and the *Thema* subject and qualifier codes for a particular book or product can be embedded within industry-standard ONIX metadata files.

Resources and further details

The *Thema* section of the EDItEUR website is at http://www.editeur.org/151/thema. This page contains links to:

- the *Thema* v1.2 codes and headings, available in multiple languages and file formats (further language versions will be made available as translations are completed)
- basic instructions for use of *Thema*
- a range of mappings (for example a procedure for mapping from BIC to *Thema*, and an automated tool for mapping from the US-focused BISAC scheme to *Thema*) for re-classification of backlist titles
- links to other resources

Further resources and links will be posted as they become available.

There is an online interactive category browser at https://ns.editeur.org/thema showing full and partial translations in around fifteen languages, including Danish, English, French, German, Hungarian, Italian, Norwegian, Polish, Spanish, Swedish, Turkish, Japanese and Arabic – Chinese and Greek are expected soon, and further translations will be added as national groups make them available)

Information about national groups and governance can be found at http://www.editeur.org/153/maintenance-and-support

A mailing list for implementation and support questions is available – sign up at http://groups.yahoo.com/group/thema_implement/join

Other enquiries about *Thema* should be send by e-mail to info@editeur.org

Graham Bell
EDItEUR, May 2016