Thema International Steering Committee Minutes

Wednesday 19th October 2016, 10:30–12:00 CEST (UTC+2),
Room Brillanz 4.2, Messe Frankfurt

Attendees:

Luc Audrain (Hachette Livre)
Detlef Bauer (Libri GmbH)
Graham Bell (EDItEUR)
Maria Börman (Bokinfo)
Francis Cave (EDItEUR consultant)
Francesca Cacciapaglia, (Informazioni Editoriali)
Fride Fosseng (Bokbasen)
Alain Fournier (BTLF)
Bente Franck-Sætervoll (Bokbasen)
Renek Mendruń (Polish national Library)
Karina Luke (BIC)
Martin Lüning (MVB / VLB)
Ivan O'Brien (Publishing Ireland)

Merve Okçuoğlu, (Turkish Publishers Association)
Diane Ouellet (BTLF)
Bruno Pacheco (APEL)
Jesús Peraita (FGEE / DILVE)
Christer Perslöv (Bokinfo)
Simonetta Pillon (Informazioni Editoriali)
Suzanne Rozario (EDItEUR)
Darren Ryan (Deanta)
Chris Saynor (EDItEUR)
Howard Willows (Nielsen), chair
Jon Windus (Nielsen)
Johanne Worsaae Petersen (Gyldendal)

Summary of decisions in this meeting:

- Jesús Peraita elected unopposed as next chair of the committee
- Minutes of previous meeting approved
- Approval of new Irish *Thema* national group
- No objections to future national extensions of language qualifiers
- No objections in principle to a supra-national group facilitated by CERLALC, should it choose to join EDItEUR
- Agreement that there is no current requirement to begin development of *Thema* 1.3
- Approval of three year extension of EDItEUR's management role

1. **Welcome and introductions**

   HW welcomed delegates to the 7th meeting of the *Thema* International Steering Committee (TISC), as the outgoing Chair of this meeting.

   GB welcomed Chris Saynor, as Alex Ingram's replacement, and introduced him to the group.

2. **Appointment of the New Chair**

   HW noted that this was the last meeting of his term as chair of the ISC, and that GB had asked for volunteers or nominations to replace him. GB reported that JP was the only nomination, and JP was therefore elected unopposed. JP will assume his role at the London Book Fair 2017.

3. **Minutes of ISC Meeting held 13th April 2016, London Book Fair, and matters arising**

   HW introduced the minutes (which as usual had been available in draft form from the EDItEUR website since soon after the London meeting). HW highlighted that the 'mission statement' had been confirmed and is available on the EDItEUR website.

HW noted that the main business of the last meeting was the impending release of v.1.2 and some issues arising from that. 1.2 was successfully released at the end of May. One topic that took a lot of discussion was the use of the term 'terroir' fiction in the English and the meeting finally settled on 'narrative theme – "sense of place"', with 'terroir' being mentioned in a scope note. Another issue raised by the UK group was the possible addition of a heading for Shakespeare but HW had noted to the UK group there was little support for this at the ISC. HW reported that the UK group was still interested in this concept, but if the members of the ISC had not changed their minds, the UK could explore a possible national extensions solution (although it was not immediately obvious how this would work) or look at the other ways this concept can be expressed.

HW noted the important issue of geographical extensions for countries which have not yet joined *Thema* but have important locations within them that existing countries using *Thema* may wish to have a code for. GB covered this in the activity report.

HW noted that the Chinese national extensions were resolved and agreed in August and that a full Chinese translation will be added. GB reported he had the full translation to hand, and publication of this is awaiting approval from the relevant Chinese government ministry. GB expects agreement within the next month or so. HW noted that there was likely to be a new set of French geographical extensions due. LA had no information on this.

HW mentioned the statement from Amazon UK that was very supportive of *Thema:* HW had recently asked for an update that could be more global or at least European, but as of this meeting had not received any further statement. HW asked that the members to keep in touch with their local Amazon representatives so that any official updates could be shared with the group.

HW said other matters arising from the minutes would be covered in the main agenda, and **the minutes were approved.**

The new Irish Thema national group was formally approved. HW officially welcomed the Irish group to the *Thema* International Steering Committee.

4. **Report on current *Thema* development work and a summary of activities since last meeting**

GB covered the main points of his Activity Report. In the period since the last meeting, EDItEUR has completed the documentation work on *Thema* 1.2, which was ratified in London and published at the end of May. The initial publication included the spreadsheet, onnline browser, the basic instructions for use and the updated executive briefing, all in English. GB noted that over last four months, there have been updates of many of the languages to translate the new codes that were added in v1.2. GB has received many updated language translations, some of them translated just the new headings, and some also improved the translations of existing headings or notes. GB reminded delegates that the improvement of translations could go on at any time, so long as the meaning of each *Thema* concept does not change. Improving the translations can only be seen as a good thing. New and improved translations mean that fewer of the headings revert back to English (the 'hub language', and the only one in which heading text is guaranteed to exist). Now there are many fewer codes that revert back, because so many of the translations have been updated.

GB noted that *Thema* is now available in 16 languages (17 or 18 if you count Chinese and Greek, which are completed but not available yet). Most of them are also available as downloadable spreadsheets, readable PDF and HTML files. EDItEUR is happy to host any of the resources that any group translates and is also happy for groups to host translated resources themselves, to ensure they are easily available for *Thema* users around the world. GB also noted that there has been a lot of updating of the online browser. Every time a brand new translation appears, it can take about half a day to spot check, to make sure the codes have not got out of line in the spreadsheets and to copy over to the master files that GB keeps and make sure that the new translations go on the web browser. GB reminded the national to use the translation templates he provides to avoid even more complex processes.

GB noted that there is an updated Arabic translation. Emad Eldeen Elakehal has done a significant amount of work to bring this up to date. GB also noted that we now have Danish, German, English, Spanish, Polish, French, Swedish, Norwegian and Italian updates. Japanese is not yet a complete translation, with only about 30% translated, but it meets the needs of the Japanese trade (subject coding within the Japanese market is at very much lower level of granularity than that afforded by *Thema*). Hungarian is new translation by Magyar Books in Print, and was added only recently to the website. Lithuanian and Portuguese are also new translations, even though there are no Lithuanian or Portuguese members of EDItEUR nor any national groups. As in Poland and some other smaller countries, if we can find some way of getting them to become members of EDItEUR, then the way to national extensions and a full voice in the future development of the standard would also be open to them. Merve Okçuoğlu has done a lot of work on the Turkish headings and the user interface text, and this will be updated on the interactive browser and the EDItEUR website immediately after the Book Fair. *Thema* has also been in use in Greek, but the translation is not yet available and GB hopes it will become available soon. GB has tested the Chinese translation with the online browser and with the process that is used for producing PDF and HTML documentation, so as soon as official permission to publish is forthcoming, the Chinese translation will also be made available.

GB noted that a number of organisations were 'screen scraping' the interactive *Thema* browser and simply downloading every single web page on the browser one by one – that's 85,000 web pages – in order to reconstruct the full *Thema* database. Some may see this as simpler than downloading the 16 or so Excel spreadsheets. This was not really a problem – but there *are* easier ways of doing it… As a result, GB made the decision to make the full database available in a convenient format. GB explained that behind the online browser there is a SQL database. GB sent this SQL database file to a couple of EDItEUR members and got feedback saying that this could save them several days of work. This also means EDItEUR can easily update the SQL file so that the organisations who want can get the latest translations rather than scrape 85,000 pages every month in order to keep up-to-date. For systems vendors who want to implement a multi-language *Thema* browser in their own systems this would be a straightforward way to go. GB has made the SQL file available on request to EDItEUR members, and he noted that so far just four organisations have requested it. If the TISC want to change that 'EDItEUR members only' policy, then it could be discussed at the meeting.

GB noted that the physical server on which the *Thema* online browser is hosted has been updated – the new hardware is significantly faster and delivers a more responsive user experience. Alongside that, the search facility within the software was updated so that it is now properly independent of accents / diacritics. For example, searching for Galapagos with a capital G and lower case g were always equivalent. Now Galápagos' with a grave accent on the 'a' and Galapagos without any accents are also equivalent. This produces better search hit rates in languages other than English.

NB some Thema users may not be aware that the online browser is optimised for use on smartphones as well as on the desktop.

Some will have noticed the reliability issues with the online browser in March just before the last meeting in London. GB reported that this was down to failures of the internet link to the EDItEUR office. This was resolved in late March, and has been reliable for the last six months.

GB noted that at the last meeting we did not have a full resolution on the question of the Chinese national extensions, but that after the meeting an agreement was reached with the Chinese group with one minor change. The national extensions were sent out on the Google group mailing list (which is specifically for TISC matters), no substantive comments came back, and after a one month comment period, the new extensions became a part of the scheme. This was the first time extensions were agreed like this: previously all national extensions had been agreed at face-to-face TISC meetings. This is a similar process to the updates to ONIX code lists, and GB suggested this was a good way to proceed for future national extension updates, when they do not coincide with updates to the scheme itself.

GB explained how the existing code order was modified to interpose a new national extension code between the existing code for Tibet and its previous parent category (*i.e.* between 1FPC, China and 1FPCT Tibet, so now Tibet is a 'child' of 1FPC-CN-N Southwest China, and 1FPC-CN-N is a child of 1FPC). This interposition of a national extension may make it simpler to deal with an upcoming revision of French geographical national extensions.

> *Note the interplation of 1FPC-CN-N Southwest China means that the code order and hierarchy cannot be derived entirely from the codes themselves. Consequently, the spreadsheets for Thema 1.2 contain an explicit 'parent' column to define the hierarchy.*

GB noted that HW had completed the first dedicated *Thema* course as part of BIC's training programme. It went well and EDItEUR hopes it can extend to other countries in various ways. HW said the training was geared to the UK industry, with a BIC to *Thema* element, but that a lot of it would translate well. GB also noted that a few organisations and national groups have been running significant training courses of their own.

GB noted that since Alex Ingram had left, the level of resourcing at EDItEUR has been reduced for the last five or six months, but now Chris Saynor has joined there will be restoration of the level of activity over the next six months.

GB reminded delegates in his report that the *Thema* browser can deal with language variants – Austrian German, Mexican Spanish or Canadian French *etc*. No translations for these language variants have been submitted to EDItEUR, but the system has been tested. A French Canadian translation could be added as soon as it were ready. The other as yet unused feature is the ability to add synonyms and keywords that will make searching the browser more effective. For future translations it may be worthwhile adding an extra column for synonyms. JP noted that synonyms might solve many of the problems with minor language versions in a very lightweight manner, for example, in Mexican Spanish where there are just a few word alternatives, such as '*computadora*' vs '*ordenador*', and synonyms for searching might be simpler than a full variant translation where the variant differs in only a few vocabulary terms. GB agreed: we can make a judgement about whether a variant should be implemented as a full variant or just by adding synonyms.

LA asked if this meant we should have some kind of semantic system so that we enrich the capacity to give answers to any requests and return the most appropriate codes. GB replied that sounded too ambitious – but there are links to the semantic web since the structure of *Thema* (and also of ONIX codelists) conforms with the structure of SKOS (Simple Knowledge Organisation System):

Thema	SKOS
Top level	topConcept
Code	notation
Heading or label	prefLabel
Note	scopeNote or editorialNote
Synonym	altLabel
Parent / child relation	broaderThan, narrowerThan
See also relation	related

GB also noted that there is has always been a plan to have canonical URIs for each heading in Thema rather than a 'dynamic dns DNS' URI, though this has been delayed because of the nature of the EDItEUR office internet connection.

GB reported that the rest of the work over the last 6 months had been around national extensions, for example working with the new Irish group which has provided an early draft of a set of proposals, and with the UK group to add more granularity to the existing UK extensions.

HW asked for any other comments or questions on the activity report.

JP mentioned interest in *Thema* within Latin American countries, and particularly CERLALC who wish to be a supra-national group to represent its member countries in Latin America. They are working on national extensions for geographical areas. JP also mentioned that for example in countries such as Mexico there are at least 30 versions of the Maya language, and there are books published about these languages. JP suggested adding 'national extensions for languages' when there is a need in a country to express that there are books *about* a language not in the existing *Thema* list. **The group agreed there was no objection** to this. [JP emphasised that *Thema* does not indicate the language *in which a book is written* – this would normally be indicated using the <Language> tag in ONIX, and there are other established methods for adding language codes to the relevant ONIX list as they are requested.]

DB noted that *Thema* was meant to be a simple scheme to allow books to be classified for sale, and that we need to be careful of overloading the national extensions with, for example, lots of language codes that will never be used. DB gave examples of Australian aboriginal languages in BIC that were never used in Germany. HW and JP replied that this was an advantage of *Thema*: you do not *have* to use the extension – it can be resolved to its parent. And it is important that *Thema* allows countries as much freedom as will fit into the structure of the scheme if there is a need within their markets. A balance needs to be struck.

GB said that *Thema*, as stated in the mission statement, is supposed to be a relatively simple subject category scheme for the book trade. But the simplicity is not about the *number of codes*. It is about the *complexity* of the scheme's overall structure and the ease of use and understanding.

JW asked for and received **confirmation from the group that they were happy to have a supra-national group, should CERLALC seek to join**. HW noted there is already the model of the 'DACH' German language group. GB said that he had told CERLALC and that it should be a principle, that if a country covered by a supra-national group wishes to create its own national group, then the country will take precedence over the supra-national group.

GB said the paper on geographical extensions for unrepresented countries should be taken as read, due to time constraint, but that it will be interesting to compare the tests that were included in the paper of Turkey and Mexico with the national extensions for these countries when they arrive. CS will be taking up the Geographic working group. GB said it would be good if groups could send in suggestions for Geographical national extensions based on the paper circulated as well as comments on the paper.

HW also said that if any group had other countries that might benefit from using their local versions of *Thema* then they should reach out. He said the UK group, for example, is keen to connect to some key export markets that are not yet members, particularly Australia, New Zealand and South Africa.

GB also drew attention to the early stage 'snapshot' of the national extensions developed by the Irish group. When the Irish group has agreed on its final proposal, then these will be circulated to the whole group.

5. Discussion of potential for further development of *Thema* 1.2 and development of *Thema* 1.3

GB noted the points above about language variants and synonyms as so far undeveloped areas of the scheme.

The general aim for *Thema* is an update of the scheme roughly every two years – though as BFS stated, this should always be dictated by real-world requirements rather than by the calendar. GB had not yet detected anything that indicated a need to start work on 1.3, and **committee members indicated they agreed**. The group should look at this question again at London Book Fair 2017 as a full revision can take 9 months at least. BFS asked if there had been many new suggested codes so far? GB replied he had received a few, but still in single figures, and none particularly 'urgent' – mostly just refinements. HW and GB emphasised if anyone had suggestions, then to send them to EDItEUR.

RM raised a question concerning 'audience' or 'treatment', and how far *Thema* should go in describing these? Should it be limited purely to subject, or should it also indicate audience interest (general, professional, academic…) or treatment (popular, serious, technical…)?

GB said it was a very good question. *Thema* does this to a certain extent already – for example you can indicate that a book about science is aimed at children, at education or vocational training, is popular science for adults or is a research level publication. Many subjects have popular, professional and research-level categories, which should be used according to the 'treatment' of the work.

RM said we should be looking at the record as a whole, which would suggest that 'audience' should not strictly be a part of a subject scheme. But because not all publishers used ONIX, this might be problematic. HW agreed with the 'whole record' approach: *Thema* was not just a subject classification, but also dealt pragmatically with the split between children's books and adult, and between different 'treatments'.

GB asked the committee whether it felt that the treatment was something we try to minimize in *Thema* so it can expressed elsewhere, or if the group should consider looking at *Thema* being more audience / treatment as well as subject. This is a big question and should be considered in the future, possibly by a specialist-working group.

LA said that he thought ONIX did not contain anything about the notion of 'treatment' and GB confirmed this was true, although it does differentiate between general adult, professional and scholarly in <Audience>. GB said the group may want to consider 'treatment' qualifiers (or combining treatment with existing interest qualifiers) when looking at 1.3 or more likely some future 2.0.

6. Updates from National Groups

HW then asked if any of the national groups had any updates, as time was limited.

AF reported that the French Canadian group had finished its translation of *Thema* v1.2. The group suggests about 150 major differences and about 1600 changes in all from the (Metropolitan) French translation. GB confirmed this would be set up as new language (rather than using synonyms). AF also said that from spring 2017, French Canadians would switch to *Thema* only, and make use of reverse mapping to legacy BISAC and CLIL.

7. Any other business

HW suggested that it would be a good idea to have a standard template for *Thema* presentations.

JW reminded the group that they had the right to appoint or dismiss EDItEUR as the 'managing agent' for *Thema* and that after three years, the agreement was up for renewal. HW called for confirmation of renewal, and **the committee was unanimous in favour of EDItEUR continuing in its current role** for a further thee year term.

8. Date and location of next meeting

HW advised that the next *Thema* ISC face-to-face meeting will be during the London Book Fair, venue and time to be confirmed. There being no other business, HW closed the meeting.

Chris Saynor / Suzanne Rosario
EDItEUR
19th October 2016