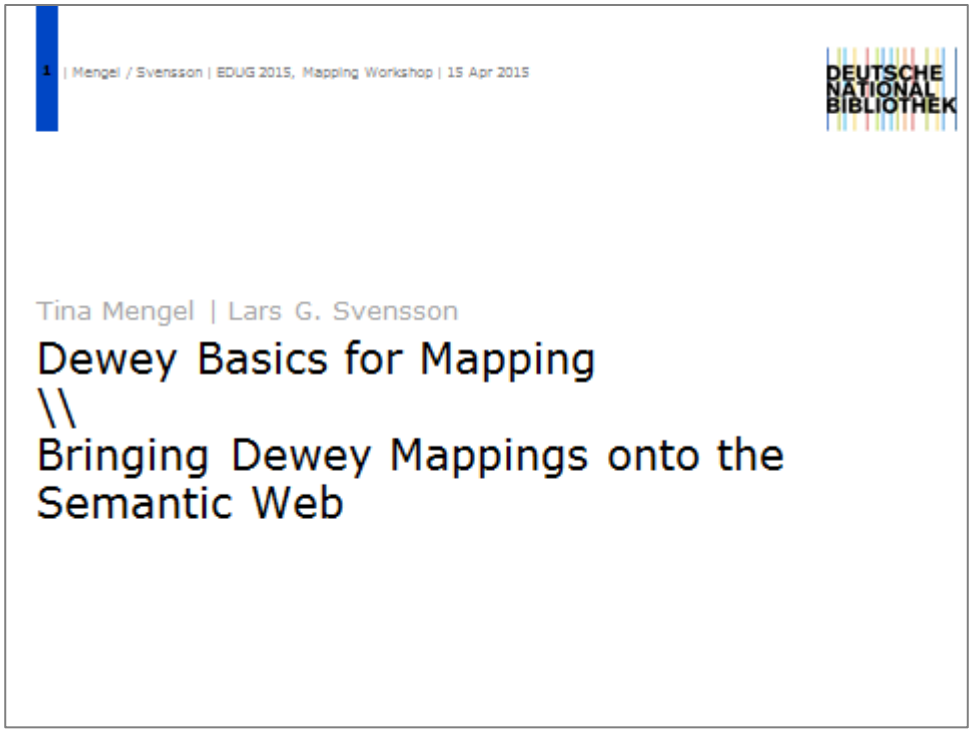
[This is an accompanying document provided by Tina Mengel and referring only to the talk “Dewey Basics for Mapping”]

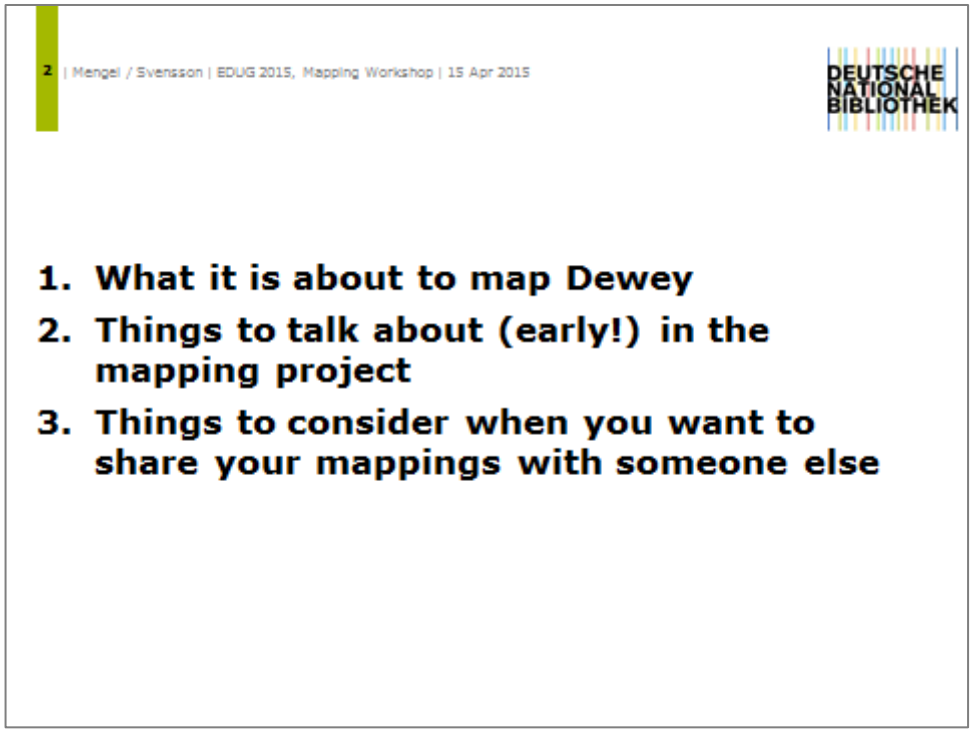


Our presentation will be divided into 3 main parts. Mine will be Top 1, which is more or less a summary of reasons why the DDC has become such a demanded scheme for mapping,

and Top 2, where I want to go into details about what to consider most when a mapping project knocks on the door.

The focus will be in the first place on the intellectual challenges and not on machine-assisted mapping.

Lars will talk about Linked Library Data and and ways to represent mapped data in general. He will discuss problems and approaches to solve the problems that come along with the representation of heterogeneous types of mappings.



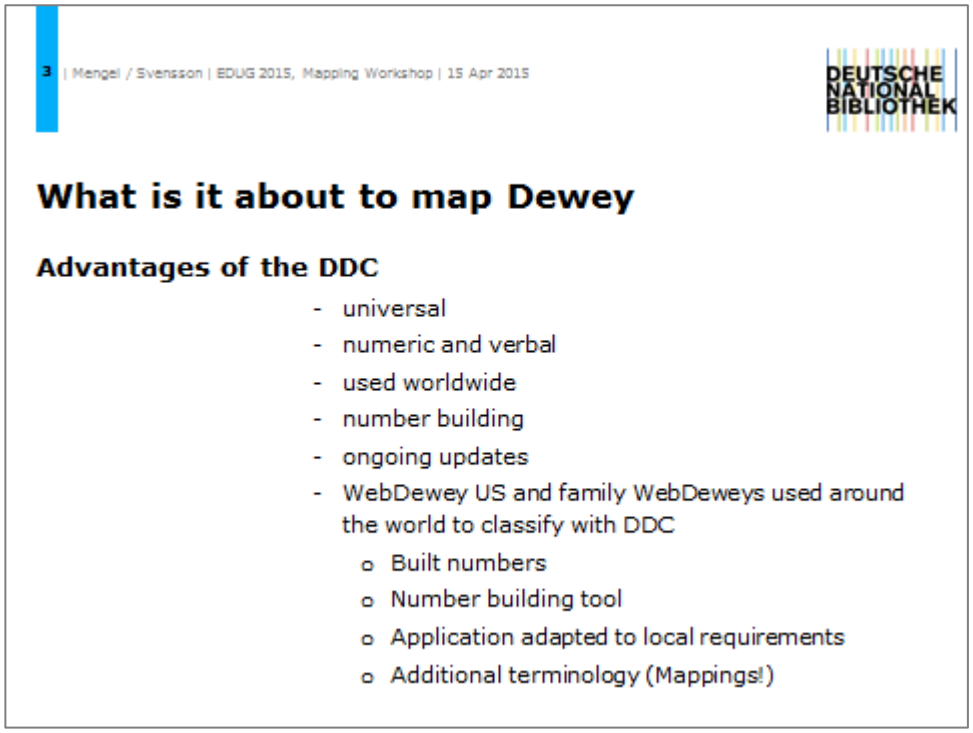
- The DDC is a universal classification system, meaning that it covers all disciplines of knowledge
- DDC numbers are language-independent and each number represents a topic or a bundle of topics, plus Relative Index terms
- The DDC is used worldwide - that is why Dewey communities meet around the globe on a regular basis to discuss DDC issues and to share experiences and projects with Dewey, and work on beneficial developments like we do today
- Dewey is permanently growing. Updated Dewey content mirrors new literature that in turn mirrors the library users' information needs.

This is what have been some of the arguments in favor of a Dewey mapping at CrissCross times. Today I want to add another argument:

WebDewey has become an internationally used tool for DDC classification and for the various working with Dewey in general. The number building tool makes it easier than ever to create new numbers for complex concepts, and:

WebDewey is the place where mapped terminology enhances verbal access to DDC classes. Not to speak of possible future developments for multilingual search and display of Dewey content and mappings..

But WebDewey is only ONE place where mappings may be represented and used. Your mapping project may still consider other scenarios to integrate mappings.

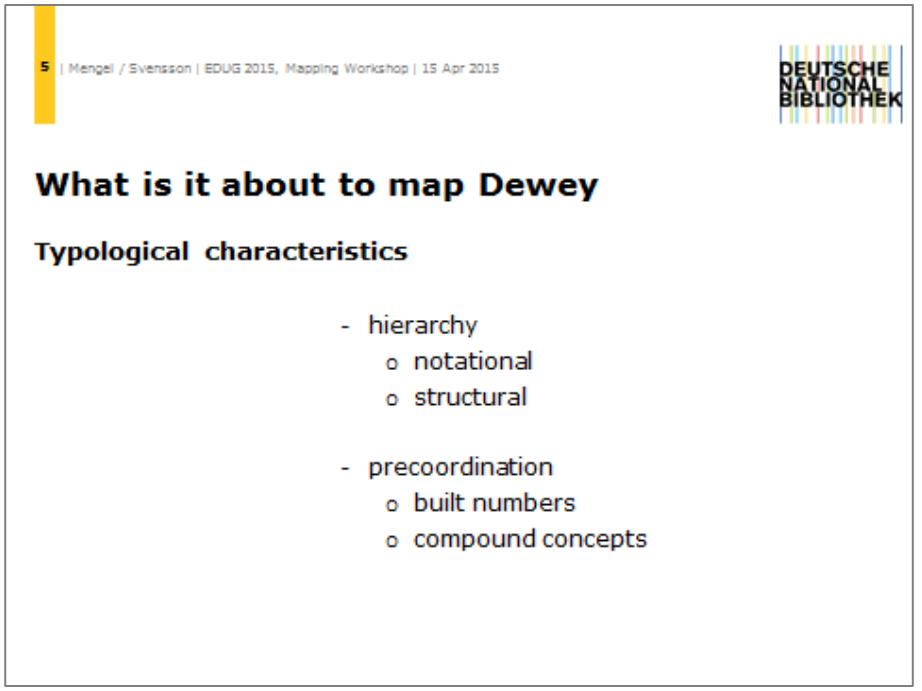


What are the typological characteristics of my system to be mapped compared to the typological characteristics of the DDC? This alone can make some decisions easy - or more complicate.

The DDC is strongly expressed by hierarchy: Notational hierarchy is expressed by the length of a Dewey number, meaning that a subordinate number usually has more digits than the broader number.

The structural hierarchy takes place more on the subject level and is represented by the logic that every subordinate topic can be seen as a part of all the broader topics above it. We will come back to that issues later.

When concepts have been combined prior to the system's use for indexing or classification, you speak of precoordinated concepts. In the DDC this is represented by built numbers, but also by combined topics in the class headings, in the notes, and in the Relative Index.

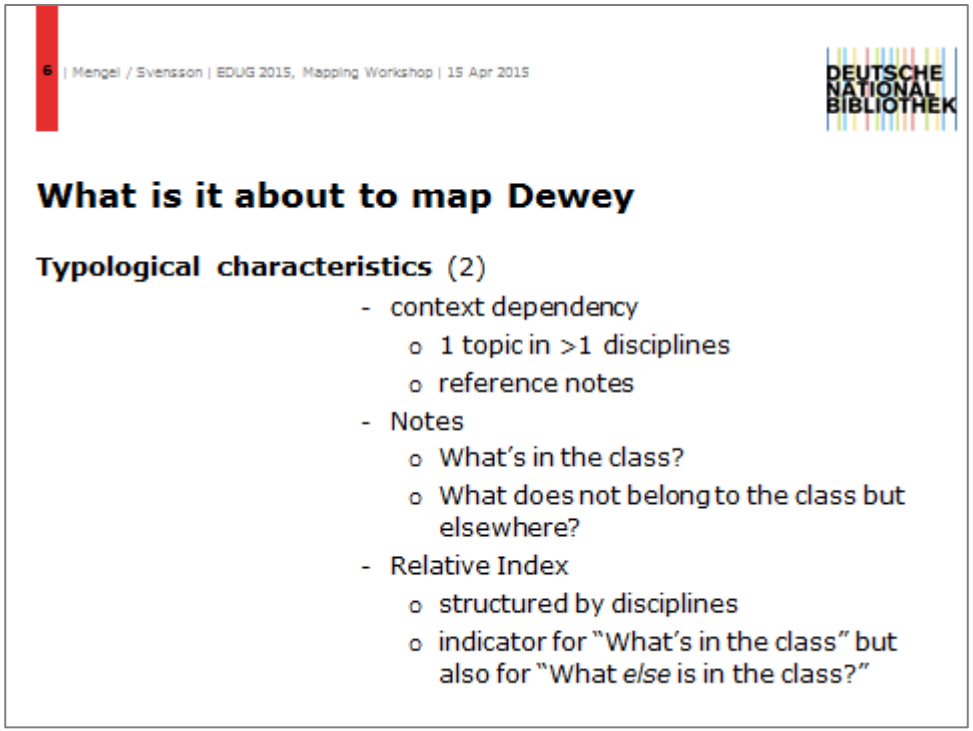


Due to the fact that the DDC is arranged by discipline, topics may appear in more than one place in the classification.

In the DDC, one topic – or specific aspects of a topic – can occur in more than one place of the classification. This is also the main characteric of the Relative Index, because there, topics are structured by disciplines. When it comes to mapping the DDC, the Relative Index is one of the most helpful thing, not only because of its representation by disciplines, but especially because of its functioning as an indicator for additional content of a class.

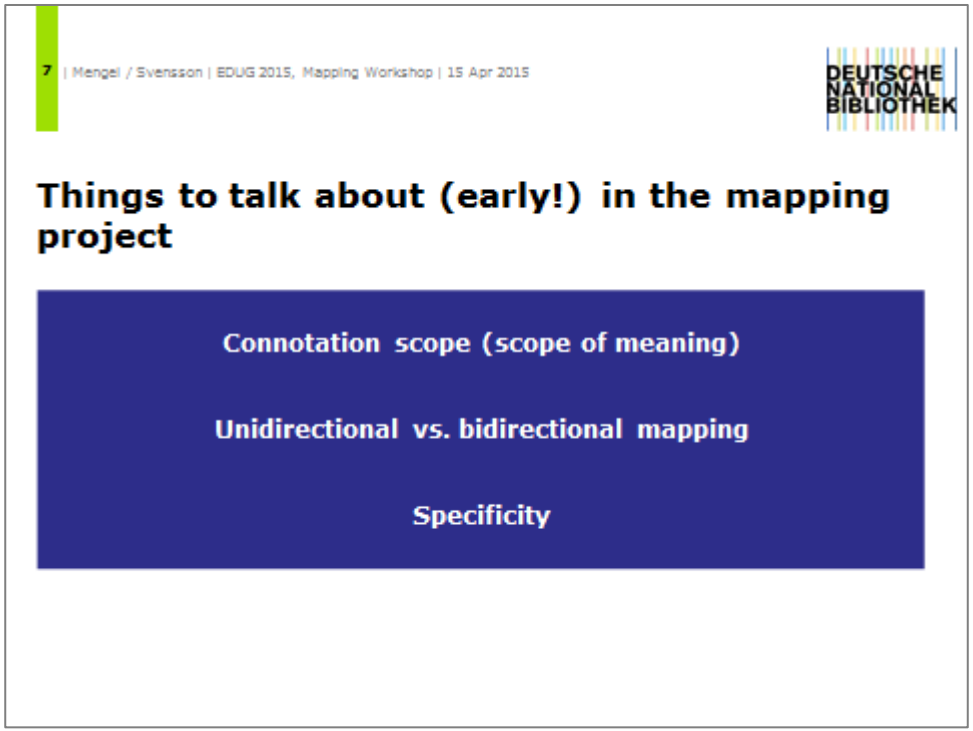
Notes are a powerful feature of the DDC. Notes tell us, what's hot and what's not in a class.

And they take our view away from the class to show us what other aspects of a topic there are in other disciplines or further up or down the hierarchy.



As we have now recalled the most important basics of DDC's typology, we can now take a closer look at the issues that really matter in a mapping project.

- Where is the beginning and the end of the entity to be mapped?
- What kind of linkages do we want to establish between the systems to be mapped?
- Up to which level of precision do we want to map?

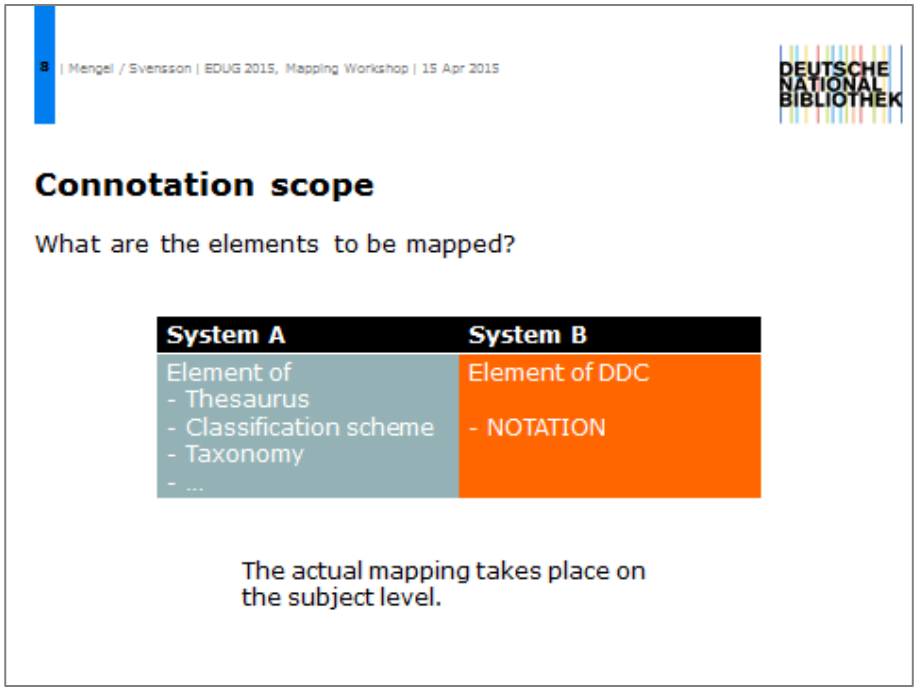


In the System A you have an element of a thesaurus, of another classification scheme, of an ontology or taxonomy whatsoever.

The target element in the DDC is a DDC number, a notation.

That's the surface.

But the actual mapping takes place on the subject level. It is about to find the best common verbal denominator, if you like.



This graph is based on the classical triangle of meaning. The principle of Concept – Object and Label is probably quite known in this room.

Winfried Gödert has summarized this to call it a Simple entity. This kind of consolidation makes sense if you want to make comparisons or statements about the interoperability of semantic elements.

Just to make that clear: even a Simple entity can be very complex in itself, but the crucial characteristic for a Simple entity is that it represents only ONE concept and is represented by a term that consists of only ONE string.

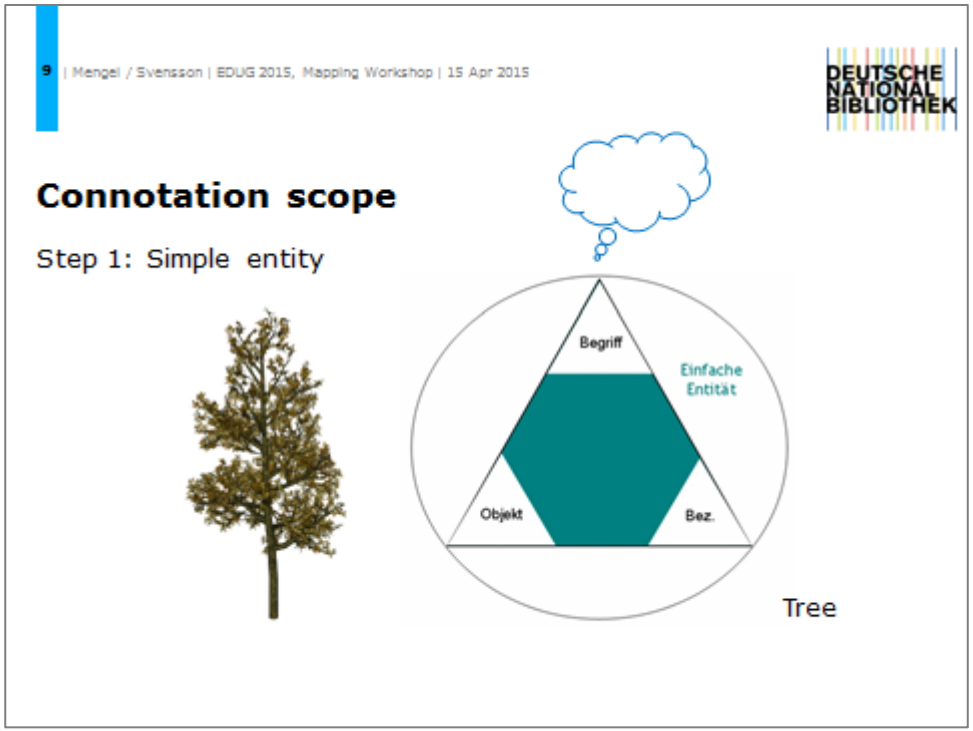


Figure (Triangle A without my poor adjustments): Gödert, Winfried: Semantische Wissensrepräsentation und Interoperabilität. In: Information Wissenschaft & Praxis 61 (2010) 1, S. 5-28

On this slide we see a second triangle: The Simple entity has become a Complex entity.

It intends to give you a notion of the high degree of semantic complexity we deal with when we are mapping elements of two schemes that are used for indexing.

Common to all kinds of controlled vocabulary that is used for indexing is that they are editorially composed and maintained collections of topical entities which are: perhaps enriched by definitions, perhaps are pre-combined or adopted to local use. They provide solution for the treatment of linguistic ambiguity problems, like for homonyms. And,

specified like this, these entities are unchangeable in meaning within the system. In the field of library and information retrieval that is an essential characteristic because it is the unambiguity of the vocabulary that makes documents retrievable.

For a mapping, however, this complexity can represent a challenge, the more, if both System A and B use different ways to handle those problems.

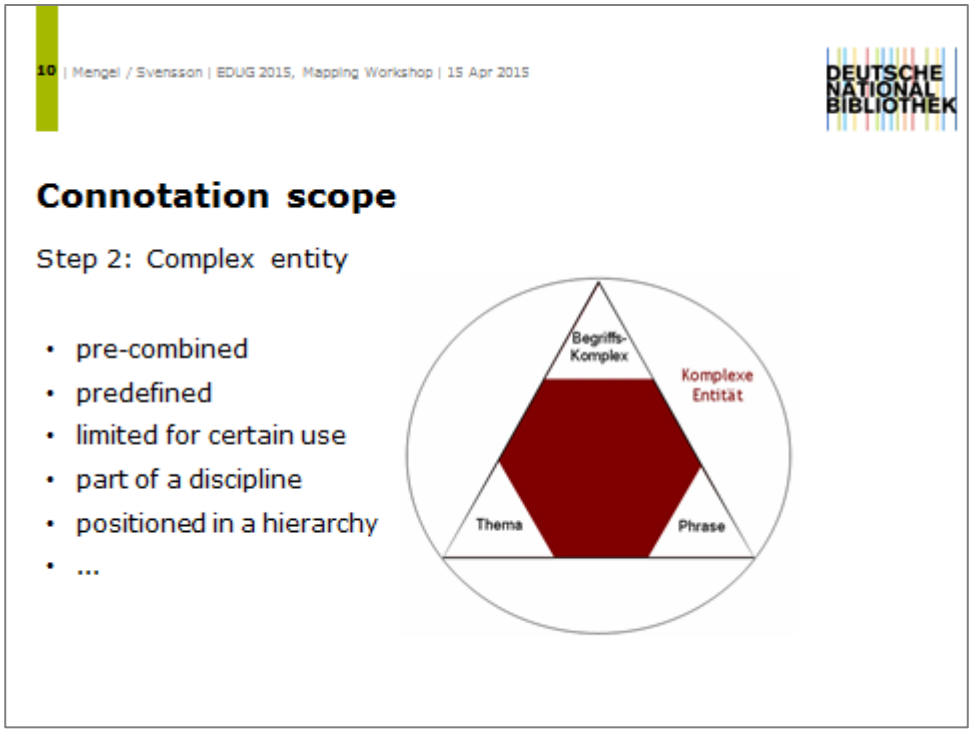


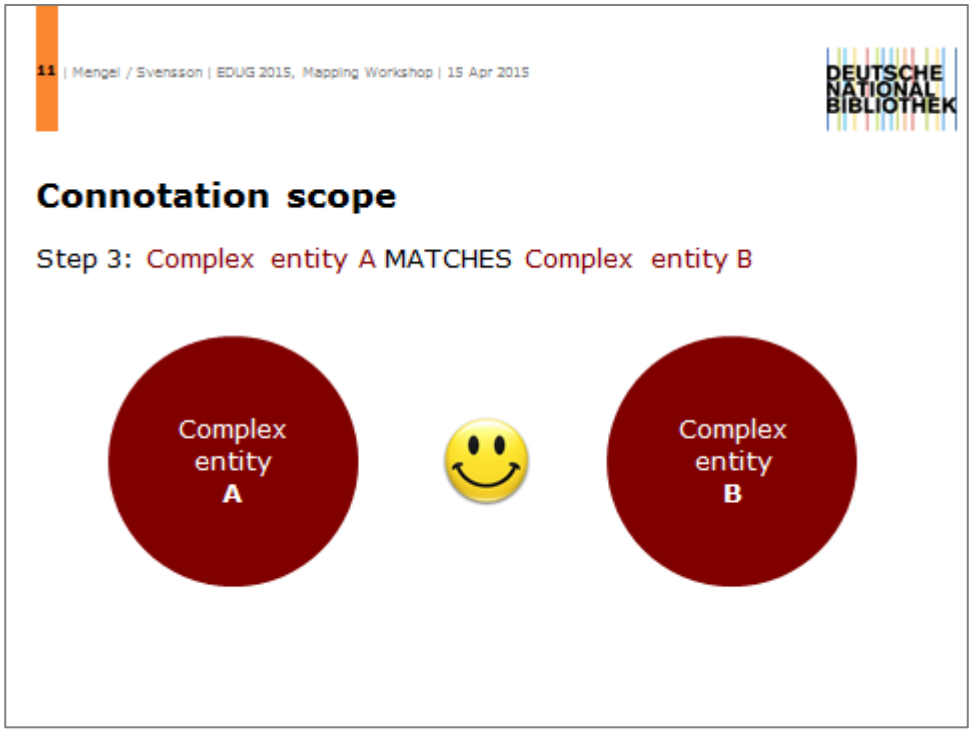
Figure: Gödert, Winfried: Semantische Wissensrepräsentation und Interoperabilität. In: Information Wissenschaft & Praxis 61 (2010) 1, S. 5-28

Well, and this is what makes mappers happy:

Complex entity A

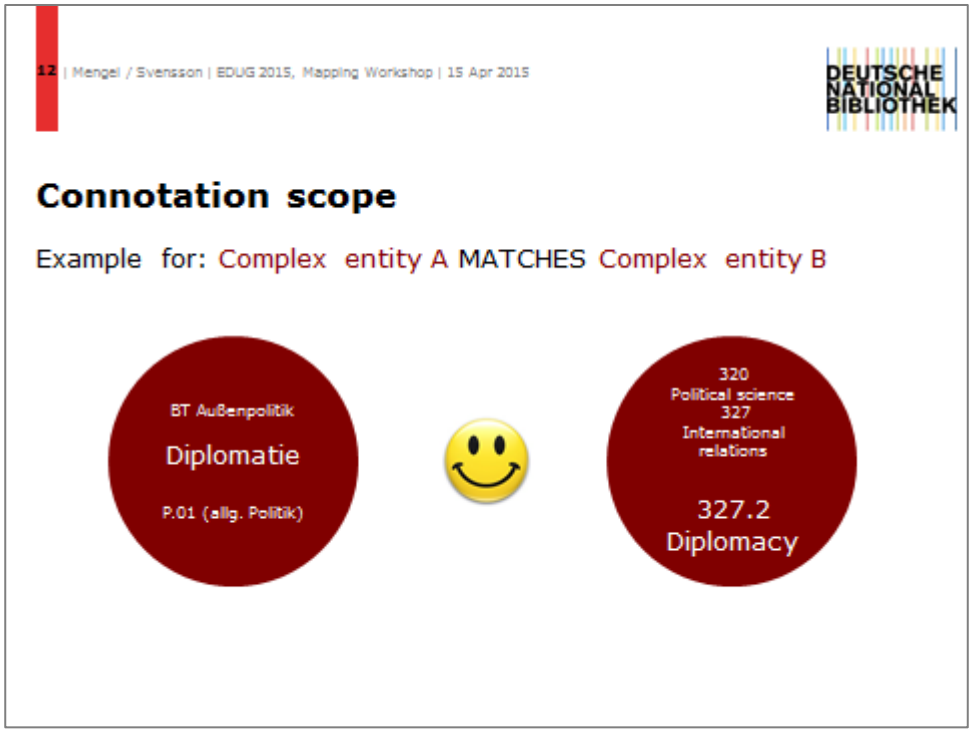
MATCHES

Complex entity B



For instance, if in the System A, like here an imagined thesaurus, the entity to be mapped is Diplomatie, it cannot be ignored that it is embedded in a relational hierarchy and in a discipline that is General politics.

This matches the DDC number for Diplomacy within Political Science in the 320s. Also here, the discipline and hierarchy and the contents of the class are part of the complexity of the entity.



When we want to map a Complex entity A of the initial system, we take it to the DDC number that seems to be good enough for a mapping.

In the next step we would have to analyze the class contents. What topics are represented by the number? Is our entity A a bingo to one of the topics in that class? And how is our entity – if found in the class – related to the other topics in the class?

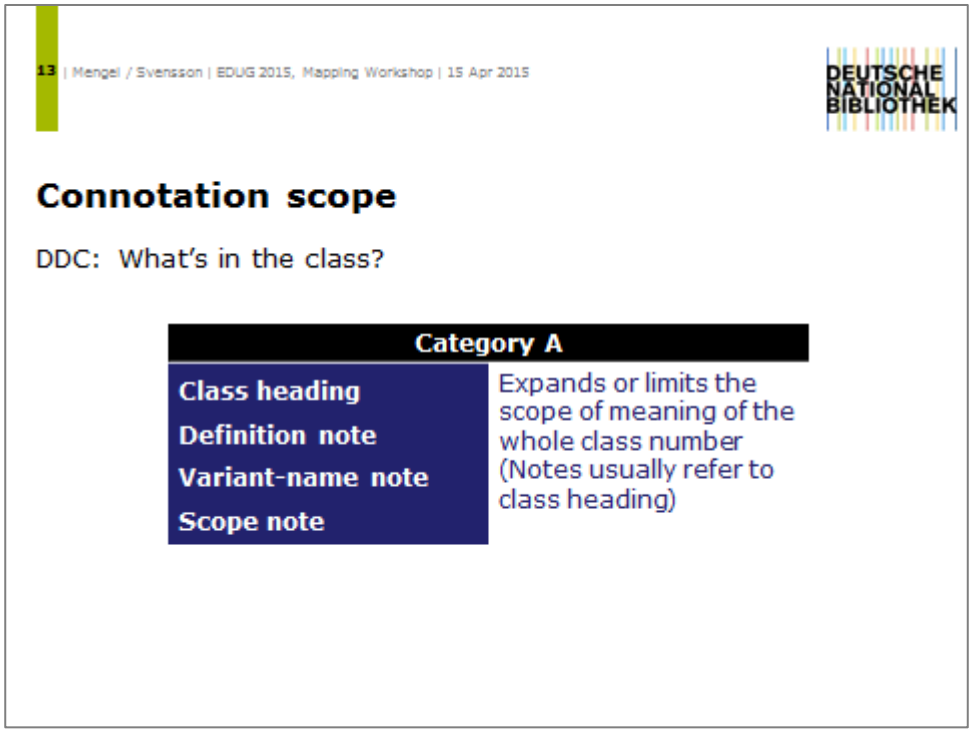
To make this a little more clear I have listed the "What's in the class" notes, and in addition, I have assigned three categories of connotation scopes, to give you an idea of how broad or narrow a topical space can expand or limit within a class.

In the first table, the connotation scopes of topics that are found in class elements of Category A affect the whole class number.

Please note that Definition notes, Variant-name notes and scope notes do not only have an impact on the present class, but may also affect subordintate classes as they have "hierarchical force".

Example:

Scope note in 155.84: Limited to ethnic groups in areas where they are not predominant



Class-here notes and Including notes have a special effect on the expansion or limitation of a topic's scope of meaning.

According to DDC terms, topics in Class-here notes approximate the whole of the class, even if broader or narrower than the class heading. They can be understood as major topics of a class. One point that is often forgotten is that topics in class-here notes have influence also on other classes, as they have so called "hierarchical force".

Topics in Including notes follow the principle of "standing room": They are without sufficient literature to have their own number, and are usually narrower in scope than the class number.

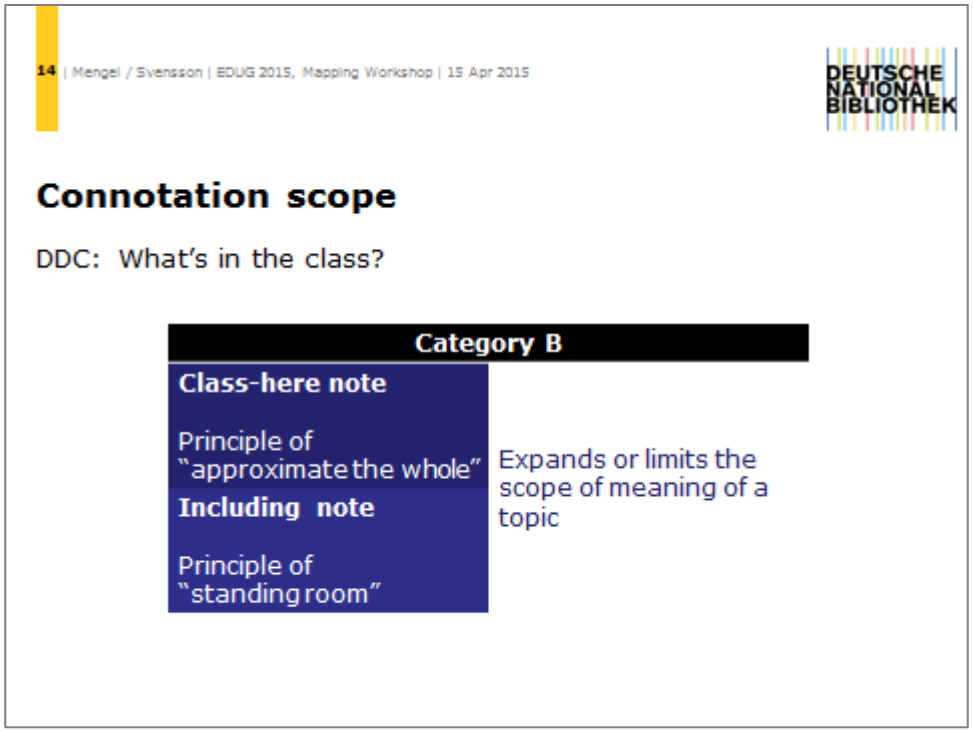
Examples:

152.1 Sensory perception

Class here: receptive processes and functions, discrimination, thresholds

152.1423 Pattern perception

Including: form perception



The third category is quite interesting because here one thing becomes really evident: both Relative Index terms and Mapped terms are not only mirroring topics that are listed in the class heading or notes, but are also representing topics that are not explicitly mentioned in the class.

Please note that the latter might pose a problem if an automated mapping approach is chosen.

Examples:

153.754 Movement perception

RI: Movement perception -> psychology

152.142 Spatial perception

RI: Space perception -> psychology -> visual perception

RI: Spatial perception -> psychology -> visual perception

Mapped GND subject headings:

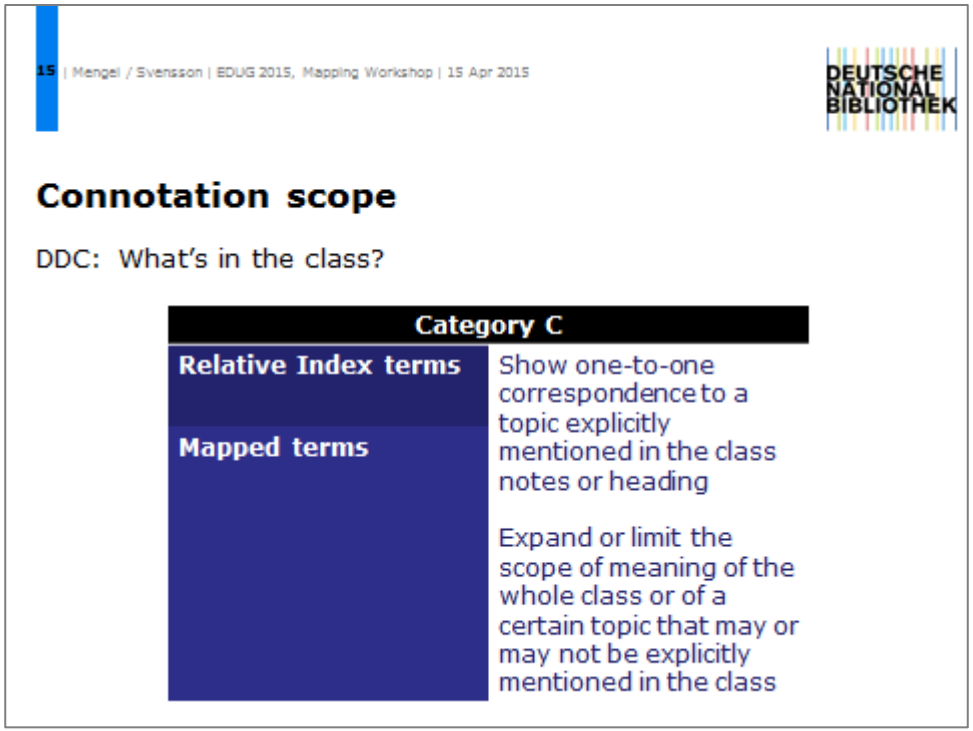
Größenkonstanz

Größenwahrnehmung

Raumwahrnehmung

Entfernungsschätzen

Raumvorstellung

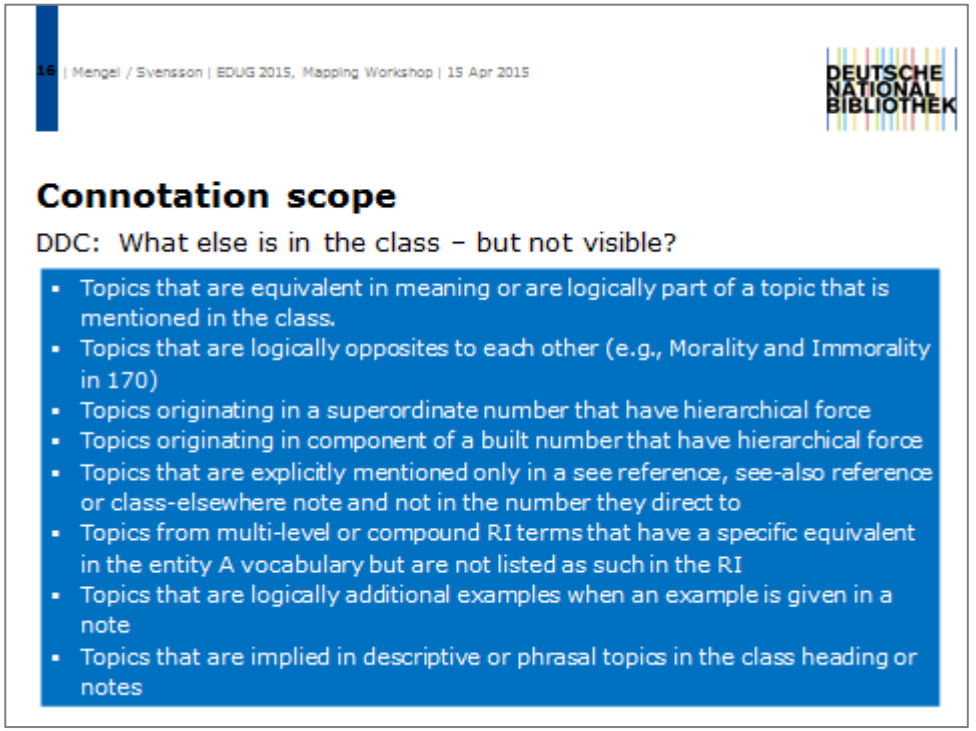


This brings me to one of the biggest questions for me in DDC mapping. How to handle those invisible contents? I only can try to give some practical advices for the time being and refer to the research that Rebecca and Michael have started at OCLC some years ago and that hopefully will be carried on.

The DDC is so well-structured and tries to be as comprehensive as possible for most of the topics treated in the classes but nevertheless, when you want to establish a comprehensive mapping to all or most entities of your initial system, you have to exhaust all possibilities of both systems.

So, what can be invisible content in the DDC?

- Topics that are equivalent in meaning or are logically part of a topic that is mentioned in the class : **Make use of synonym and hierarchical relationships in the initial system**
- Topics that are logically opposites to each other (e.g., Morality and Immorality in 170) : **Make use of hierarchical relationships like broader or narrower terms, depending how your system deals with opposites**
- Topics originating in a superordinate number that have hierarchical force : **Here I see something that the DDC could do for mappers: indicate topics that have hierarchical force, indicate subordinate numbers that are affected by hierarchical force. Up to now, mappers need to click up and down the hierarchy to recognize topics in subordinate numbers that are existing, but not visible**
- Topics originating in a component of a built number that have hierarchical force : **Same for this one: any signal can help, especially when a deep-level mapping is applied**
- Topics that are explicitly mentioned only in a see reference, see-also reference or class-elsewhere note and not in the number they direct to : **Mappers (and also Dewey translators, btw) face two problems here: Terms may be inflected or "described" and therefore are not literally identical with the term in number they direct to what makes them often hard to find.**



DDC and DDC translations can help by terminological consistency, but also automated ways or better search options for interconnected numbers to reveal the whole topical neighborhood would be so good to have

- Topics from multi-level or compound RI terms that have a specific equivalent in the entity A vocabulary but are not listed as such in the RI : **Also a tricky task for Mappers. Problems occur on both sides. One is that controlled vocabularies tend to use a lot of outdated terms and are often not in line regarding the modernness of their terms. Another problem may pose a different degree of pre-combination of the entities. Artificial term constructions or permutations may also pose a problem.**
- Topics that are logically additional examples when an example is given in a note : **Typical example: *Arrange alphabetically according to, e.g.* example - That's a nice solution for the DDC, but can cause overtime work for the Mapper who probably needs to do some exhaustive searches, depending how good relationships are modeled in the initial system.**
- Topics that are implied in descriptive or phrasal topics in the class heading or notes : **Typical formulations here are terms starting with "all kinds of" or numbers without notes or Relative Index terms, but with a Standard subdivisions heading, or combined phrases like "description, critical appraisal of"**

For all these cases, mappings are often hard to find, but once established, especially for the latter three points listed here, the value of mapped terminology becomes quite clear, I think.

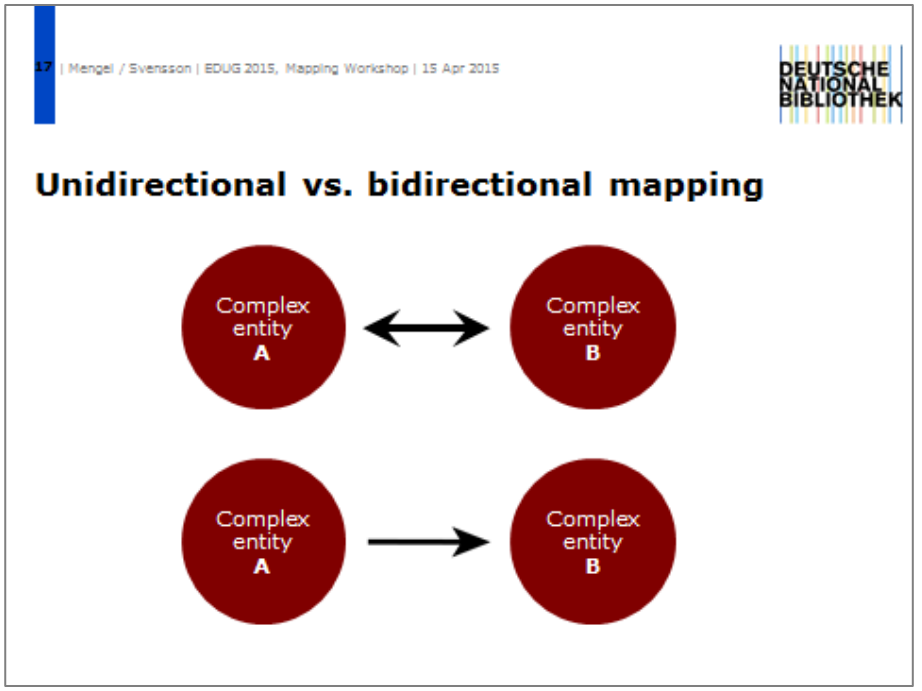
Directionality in Mappings

Questions to ask related to the directedness of mappings:

Do you want the linkages to be fully equivalent? Or do you see the bigger benefit in a directed mapping, that is, Complex entity A is contained in Complex entity B?

If the systems to be mapped are typologically similar, a bidirectional mapping might be a good solution because you can expect a high degree of semantic exchangeability of the concepts.

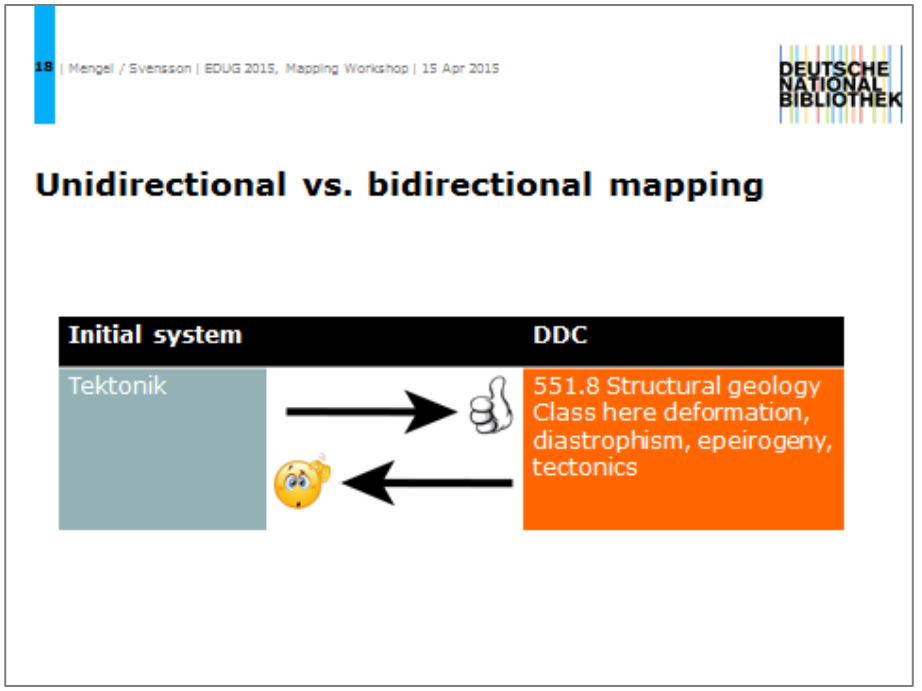
If the systems to be mapped are typologically different, like for example a thesaurus and a classification system, you may consider to chose a unidirectional mapping, because you can expect only few fully equivalents among the linked entities.



Simple example to illlustrate the problem:

Tektonik is an entity in the initial system. The scope of meaning covers only the specific topic, maybe there are some synonyms, maybe the term is embedded in a relational hierarchy. Anyways, in the end, the entity to be mapped does not cover the full meaning of DDC number 551.8 ; it is listed in the Class-here note, so it covers at least more than half of the meaning of the class.

Due to this analysis it is not possible to map the DDC number back to the thesaurus entity. The scopes of meaning are still too different, the DDC number just “eats up” the single topic.



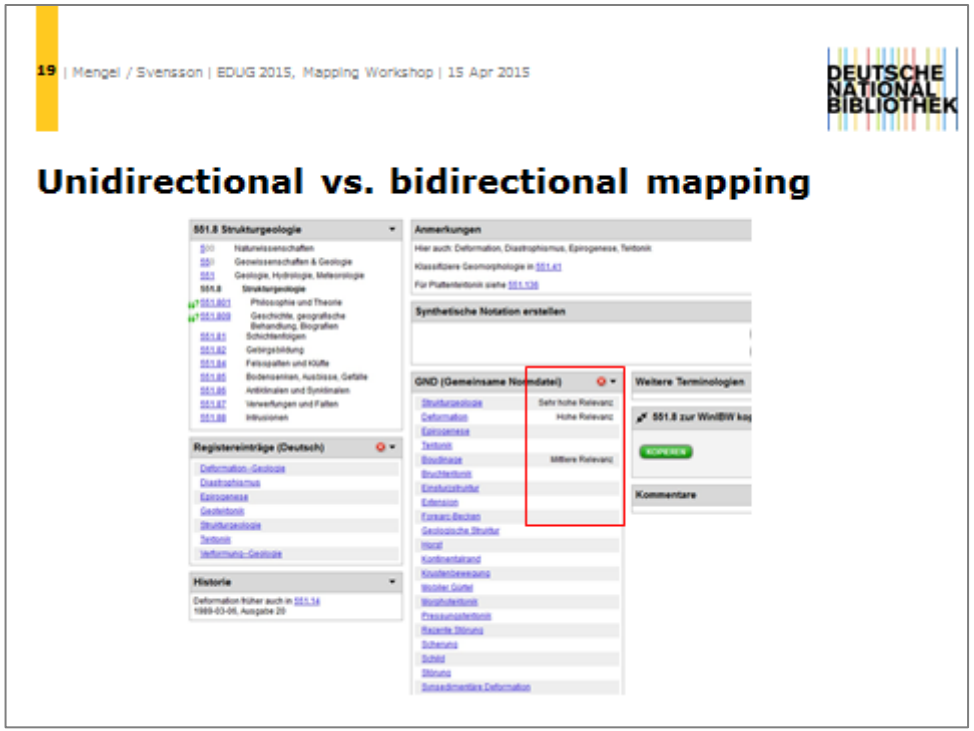
On the other hand there is no need to be sad for the single topic Tekonik, because, when we look at this slide, it is in good company:

Other specific entities of the initial system will be mapped to this DDC number, too. Any of them, together with their synonyms and near synonyms, are able to expand the connotation scope of the DDC number.

That was why in the CrissCross project we saw a great benefit in enriching the vocabulary of the target system, and by that to enhance access points to DDC-indexed resources.

When comparing the connotation scopes of GND headings and DDC numbers it became clear that we needed an instrument to indicate the relationship between the two entities, but we wanted also take into account the one-way direction of the mapping. So we developed the **Degrees of Determinacy** to express “how good our entity A fits into the DDC class”.

You can see the relevance ranking in this number for Structual geology.



In the beginning there were only three Degrees of Determinacy,

D1 for when an entity A has only a little overlapping with the connotation scope of the DDC class

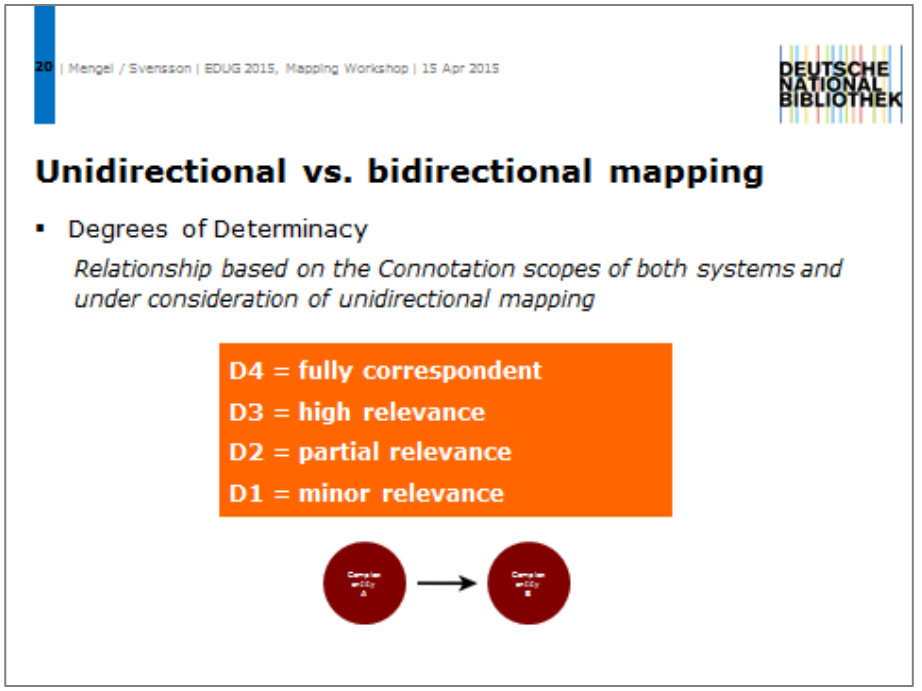
D2 for when entity A is logically a part of the class

and

D3 for when entity A covers more than the half of the meaning of a class.

But also in our one-direction approach of mapping from a thesaurus to the DDC we encountered mappings that were fully identical in meaning and which could therefore be considered as an 1:1 mapping that could also be read bidirectionally.

So, in the course of the mapping project we added a fourth degree of determinacy to make explicit that in these cases the highest possible degree of conceptual congruence of the two mapped entities are given.



Besides their function as relationships under consideration of unidirectional linking, the degrees of determinacy are also intended to serve as an instrument for the design of search options or for ranking algorithms in end user retrieval scenarios.

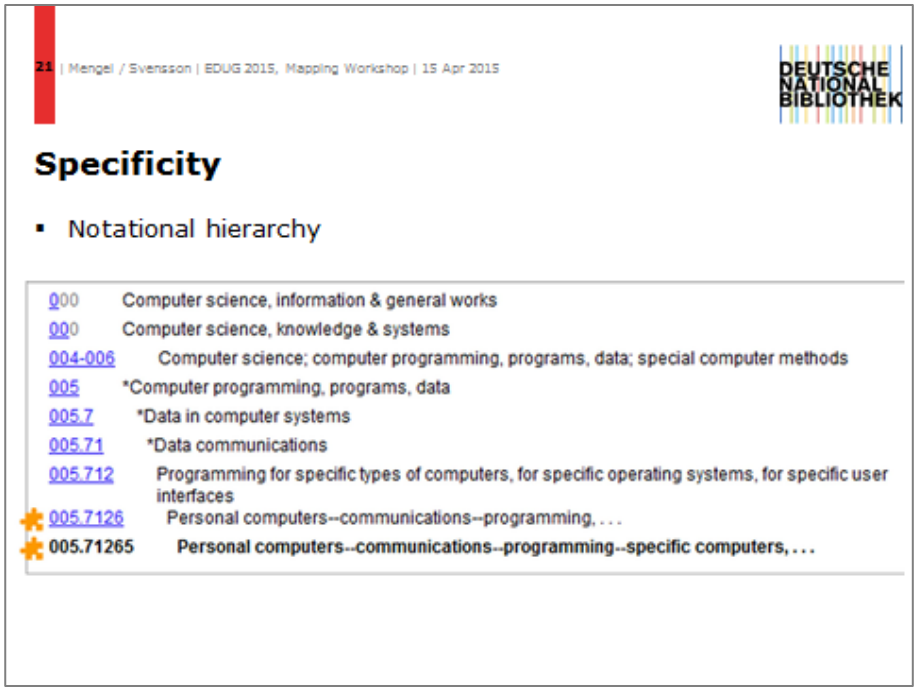
Specificity

- To which degree of precision do you want to map your vocabulary?
- What are the consequences if mapped only on the top levels of the DDC?
- How strict do you want to handle context dependency, what are the consequences for your retrieval scenario?

I cannot answer these questions for you because this requires a deep analysis of your local system to be mapped and of your general objectives of the project, for example end user retrieval environment vs. an expert concordance database.

What I can do, is to show how the DDC handles Specificity and what might be arguments in favor of a certain degree of specifity in mapping.

What has been earlier mentioned is the DDC's main characteristic of Notational hierarchy. As a rule of thumb you can say that any subordinate class is a specification of the number that is one digit shorter, depending on what has been added by number building.



So why not using this feature to get to more exact mappings?

Subject headings that are used for indexing for example can be very specific and narrow in their scope of meaning.

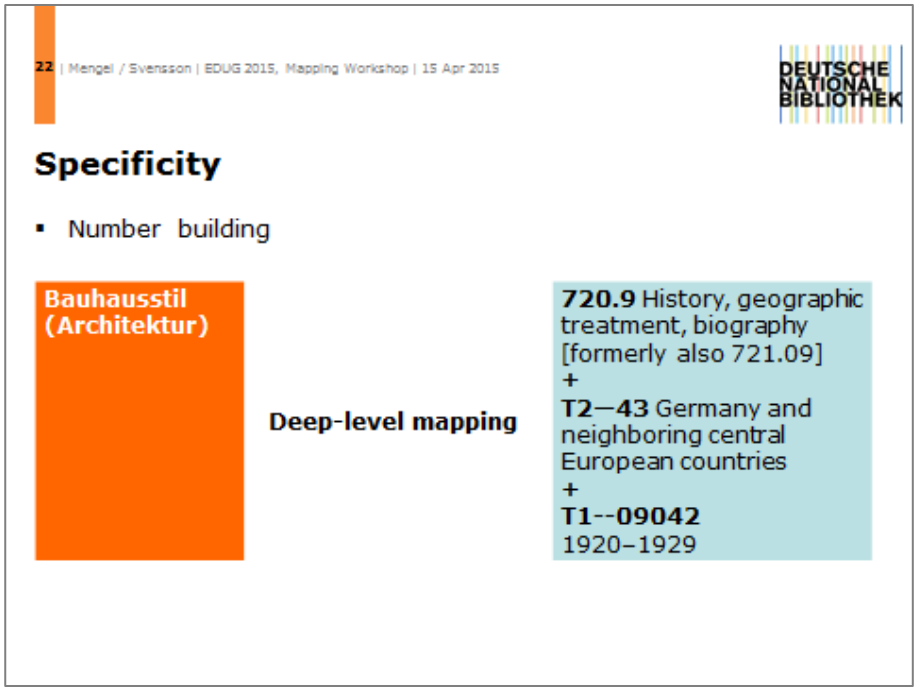
If all specific topics were put in broader classes, users searching for documents on the specific topic would probably have to deal with huge result sets containing also all the other topics of the broader class.

In CrissCross we therefore decided in favor of a Deep-level mapping, which means that we were trying to find the most specific DDC number available in the system, but additionally we decided to create numbers according to DDC number building rules if a subject heading could be represented more precisely by that built number.

In our times we still used MelvilClass for creating instituonal built numbers in order to match a specific subject heading.

Today, WebDewey provides the best platform for user-created numbers, and with the number building tool it is becoming easier than ever to produce new numbers.

The functionality of user-contribution will be developed in the near future, which will be another argument for a Deep-level mapping, because then it will be possible for all of us to share new numbers – and that will be a great enrichment also for the global DDC database.



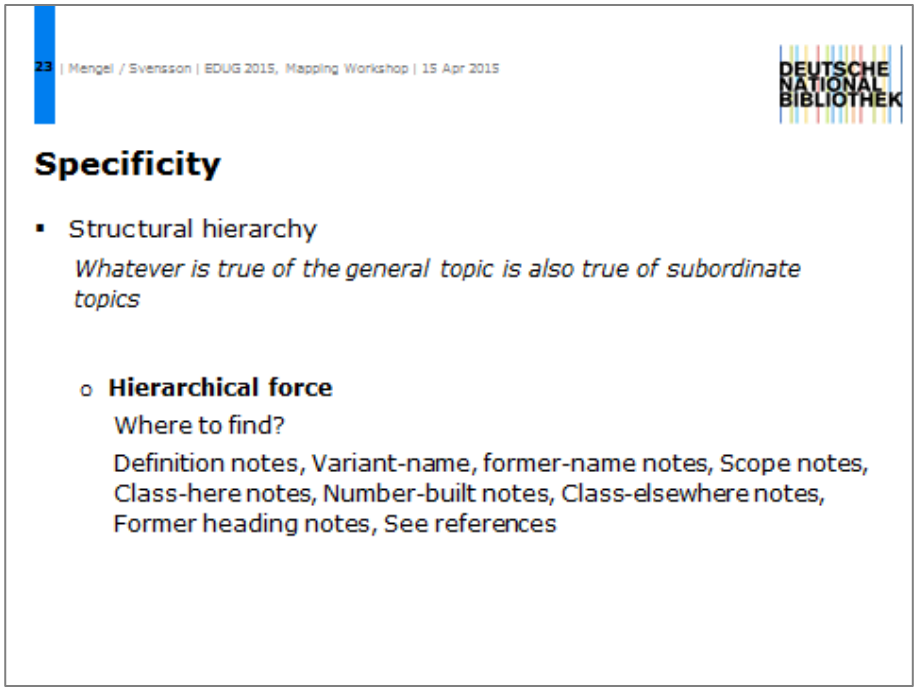
Another aspect of Specificity in the DDC is the Structural hierarchy.

I want to treat this only in a nutshell but wanted to not miss it to mention, because this is an aspect that can be easily overlooked when searching the suitable number.

Whatever is true of the general topic is also true of subordinate topics.

For the intellectual mapping process this is a rule worth a tattoo.

I remember that we had a lot of struggle with this during CrissCross because there were only few examples and explanations that time to be found in the internet and literature.



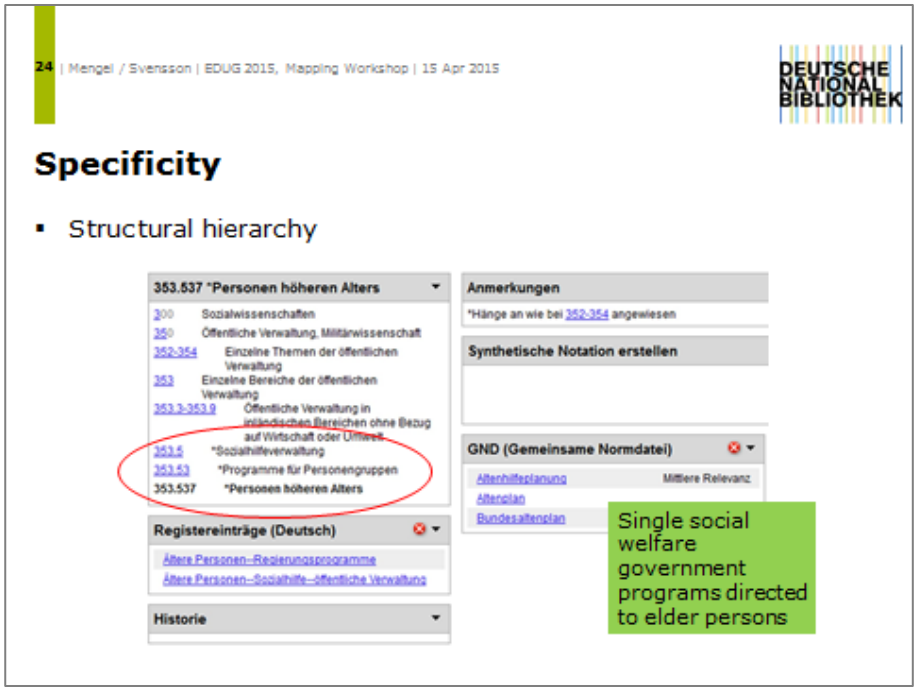
Not all mappers are also DDC classifiers – I wasn't, by the way, - and in the beginning, DDC interpretation might not be so easy.

So, again:

What is true of the general topic is also true of subordinate topics.

Without the Relative Index term in this class for Elder persons you wouldn't have any cue to understanding the contents of the class, if you didn't read down the structural hierarchy, to get to:

Single social welfare government programs directed to elder persons.



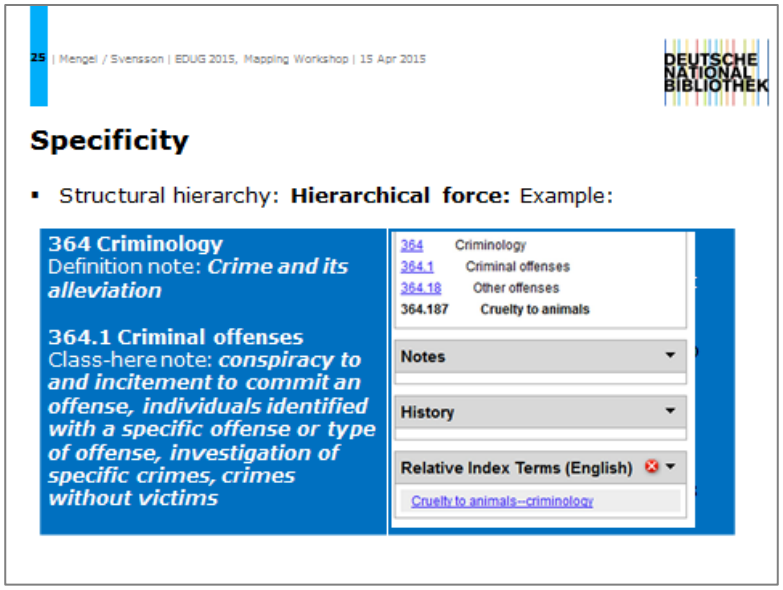
The second example is one for the principle of **Hierarchical force** in the DDC.

The number on the right side for Cruelty to animals looks quite empty. There are no notes and there is only one Relative Index term telling us nothing that we don't know yet.

If you mappers come across such a class you should hear a bell ring.

In the superordinate numbers there are some notes to be found that give us a hint of the hidden class content for the Cruelty to animals number.

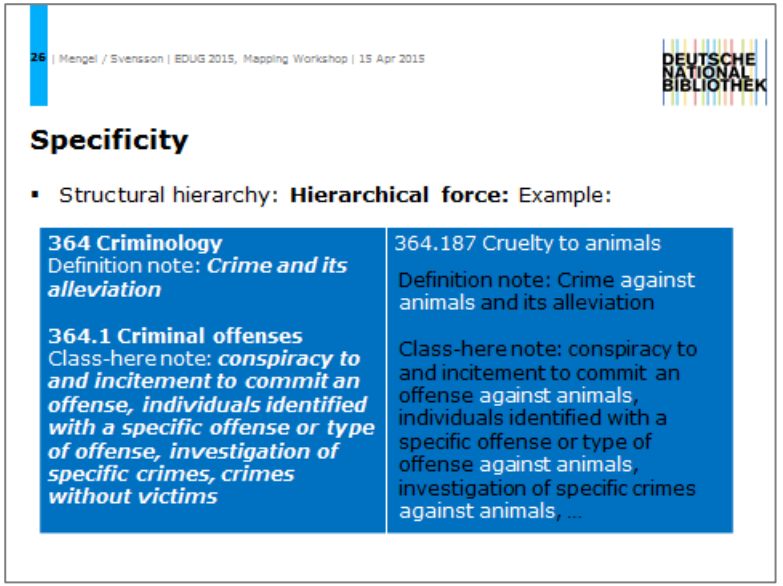
The topics in the Definition note in 364 as well as in the Class-here note in the subclass of 364 have hierarchical force also for 364.187.



Noticing that, the Cruelty to animals number is no longer that empty!

All the aspects of the notes that have hierarchical force in the broader numbers also apply for Cruelty to animals.

This gives a lot more specification to the class contents – and hence more options for mappings – than could be expected without consideration of hierarchical force.



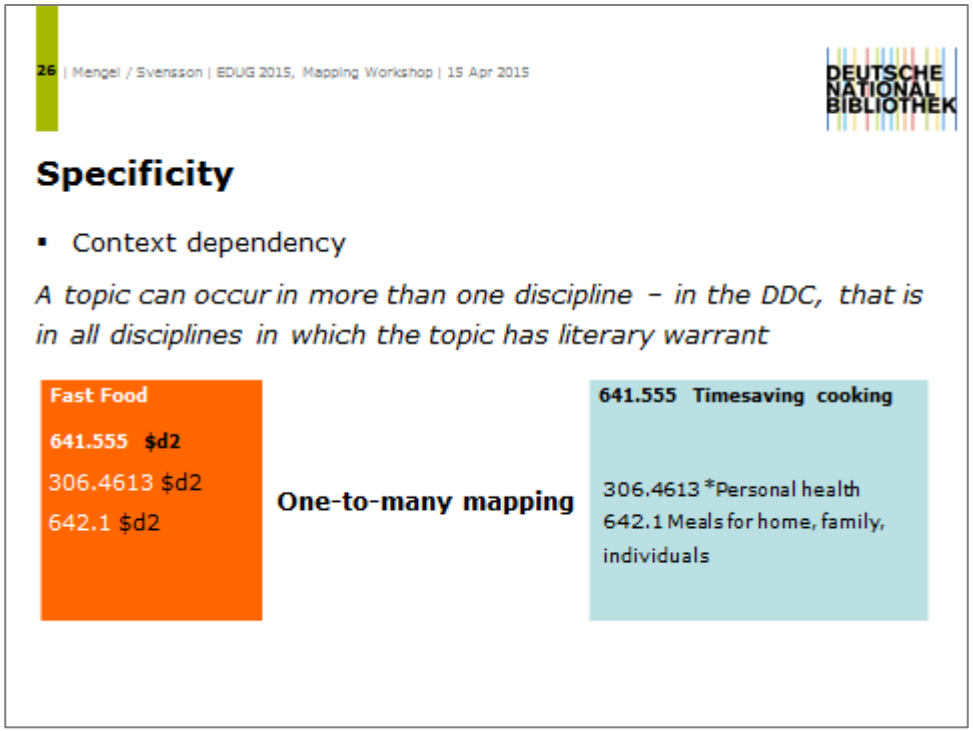
Coming back once again to the typological differences of the systems to be mapped, context dependency might be another factor that requires some attention.

In the DDC, a topic can occur in more than one discipline. Depending on what your project's aim is, it might make sense to consider mapping in any of the disciplines that your users will find helpful for their tasks.

Example:

An entity A Fast food could for example be mapped to DDC class for Timesaving cooking.

But users may also be interested in literature on other aspects of Fast food, like the Sociology of eating, or in the context of Meals and table service.



DEUTSCHE NATIONAL BIBLIOTHEK

Further things to talk about (early!) in the mapping project

- Mapping of Tables – Yes or no?
- Updating (hence versioning) of mappings according to DDC changes – Yes or no? And if Yes, how?

DEUTSCHE NATIONAL BIBLIOTHEK

Mille grazie, ragazzi e ragazze !

References:

- https://www.oclc.org/content/dam/oclc/dewey/resources/teachingsite/courses/Technical_Introduction.pdf
- http://www.oclc.org/support/documentation/glossary/dewey.en.html
- Green, R.; Panzer, M.: The Ontological Character of Classes in the Dewey Decimal Classification. In: Paradigms and conceptual systems in KO: proc. Eleventh int. ISKO conference, Rome, 23-26 February 2010.
- CrissCross project page: http://ixtrieve.fh-koeln.de/crisscross/index_en.html
- Gödert, W., Hubrich, J., Nagelschmidt, M.: Semantic Knowledge Representation for Information Retrieval. De Gruyter Saur, 2014.
- Illustrations on slides 9 and 10: Gödert, Winfried: Semantische Wissensrepräsentation und Interoperabilität. In: Information Wissenschaft & Praxis 61 (2010) 1, S. 5-28
- WebDewey US: http://dewey.org/webdewey/login/login.html
- WebDewey Deutsch: http://deweyde.pansoft.de/