



Concepts and Relations

Lukas Koster
Library of the University of Amsterdam

--

<http://commonplace.net>

2009

It started with:


A screenshot of a Twitter interface. At the top, the Twitter logo is visible in blue. Below it, a navigation bar contains links for Home, Profile, Find People, Settings, Help, and Sign out. The main content is a tweet from the user 'hochstenbach'. The tweet text reads: "Thinking about @lukask's 'RDF' query, I would like to do a #synonym search in #solr without reindexing..doesn't work for multiple words". The phrase "#synonym search" is highlighted with a red rectangular box. To the right of the text are icons for a star and a reply. Below the text, the time "4:57 PM Aug 21st from TweetDeck" is shown. At the bottom left of the tweet is the user's profile picture, a cartoon character, and the username "hochstenbach".

Twitter

Home Profile Find People Settings Help Sign out

Thinking about @lukask's "RDF" query, I would like to do a #synonym search in #solr without reindexing..doesn't work for multiple words

4:57 PM Aug 21st from TweetDeck

 hochstenbach

It started with:



The screenshot shows the Twitter profile of 'lukask'. The profile picture is a man's face. The name 'lukask' is displayed in large black text. Below the name, it says 'lukask's Favorites'. A list of tweets is shown, with one tweet highlighted by a red box:

hochstenbach @lukask Karen Coyle's problem is related in the sense that searching "War and Peace" should give you "Война и мир"
8:52 PM Aug 21st from TweetDeck

The screenshot shows a conversation thread on Twitter. The tweets are as follows:

- hochstenbach** @lukask Or a nice presentation at #datasalon <http://bit.ly/xg5Qu> ..he he
8:41 PM Aug 21st from TweetDeck
- lukask** @hochstenbach btw did you see this? <http://bit.ly/1ackxD> More or less the same problem
8:40 PM Aug 21st from TweetDeck
- lukask** @hochstenbach OK, I will find our documentation, but some time next week? #solr
8:39 PM Aug 21st from TweetDeck
- hochstenbach** @lukask could give some inspiration to the #solr crowd what libraries really want
8:38 PM Aug 21st from TweetDeck
- lukask** @hochstenbach Well, that would take some time. It was Java classes, implemented in Java servlets and MySQL #solr
8:35 PM Aug 21st from TweetDeck
- hochstenbach** @lukask Send me the source. The #solr solution doesn't work the way I want :-)
8:33 PM Aug 21st from TweetDeck
- lukask** @hochstenbach @edsu Well, absolutely! That is exactly 1 of the features of a system that me and a colleague designed and piloted 7 years ago
8:31 PM Aug 21st from TweetDeck
- hochstenbach** @lukask Oh, #solr allows synonym lists at indexation and at query time. I guess this would be a form of @edsu better cataloging solution
8:29 PM Aug 21st from TweetDeck
- lukask** @hochstenbach Tell me more? #synonym #solr
8:21 PM Aug 21st from TweetDeck
- hochstenbach** Thinking about @lukask's "RDF" query, I would like to do a #synonym search in #solr without reindexing..doesn't work for multiple words
8:17 PM Aug 21st from TweetDeck

In the mean time ...



Coyle's InFormation

Comments on the digital age, which, as we all know, is 42.

SUNDAY, AUGUST 16, 2009

➤ What is a (FRBR) Work?

"What is a Work?" is an oft-discussed question. Answers tend to come down on one side or another of what is essentially a philosophical reaction to the inherent abstractness of the nature of the FRBR Work.

I was poking around in the Futurelib wiki (much neglected of late, but I have recently [gathered there](#) all of my posts on Martha Yee's article on library cataloging and RDF), and came across an [interesting comment by Kristin Antelman](#) from over 2 years ago:

I realize this is a point of absolutely no controversy in the FRBR community, but I have never been happy with the title attribute in association with the abstract entities, work and expression. It seems contrary to the spirit of an abstract entity, not to mention creating practical problems (e.g., for serials). There obviously could be many titles associated a work in its manifestations. Libraries may want to select one over another for the "work," or display, title. Works and expressions only need identifier attributes: for the work, author and subject.

RDA recognizes both a Work title and a Manifestation title (called title proper). I've been known to argue that these are distinct data elements: the Manifestation title is transcribed from the piece and is

Karen



Karen Coyle

Labels

- [application profiles](#) (1)
- [authority control](#) (4)
- [books](#) (3)
- [cataloging](#) (19)
- [classification LCSH](#) (3)
- [copyright](#) (9)
- [Digital libraries](#) (9)



In the mean time ...

RDA recognizes both a Work title and a Manifestation title (called title proper). I've been known to argue that these are distinct data elements: the Manifestation title is transcribed from the piece and is part of the surrogate for the manifestation; the Work title (which we used to call "Uniform title") acts as a unifier for all of the Expressions and Manifestations of the Work. Kristin's comment got me thinking about this again, and I agree with her: there is no title for the Work. The Uniform title served as an identifier for the Work in the days when we used things like titles to identify things. But FRBR and RDA recognize that entities will have identifiers that are separate from the display forms of names and titles that we have used in the past.

This "no title" solution actually helps out with one of the sticky problems in creating Work displays, particularly in a multi-lingual catalog. When you follow the concept of uniform titles, the Work title should be the title of the original. This means that we would be showing our users **Война и мир** as the title for the Work that most of them will know as *War and Peace*. We could show them the English language title, but what if your catalog users are global? What if some of them will only understand the title if you display it in French or Turkish or Chinese? If a Work has an identifier (which is only useful for machine processing, not for display to humans), then you can let users choose what language they prefer in Work displays. (Obviously having some default for the case where the user's preferred language isn't available.)

Abstract identifiers

Multilingual labels

So, the issue is:



■ Objects

- Abstract unique identifiers
- Synonyms
- Multilingual labels

Once upon a time in a galaxy far far away

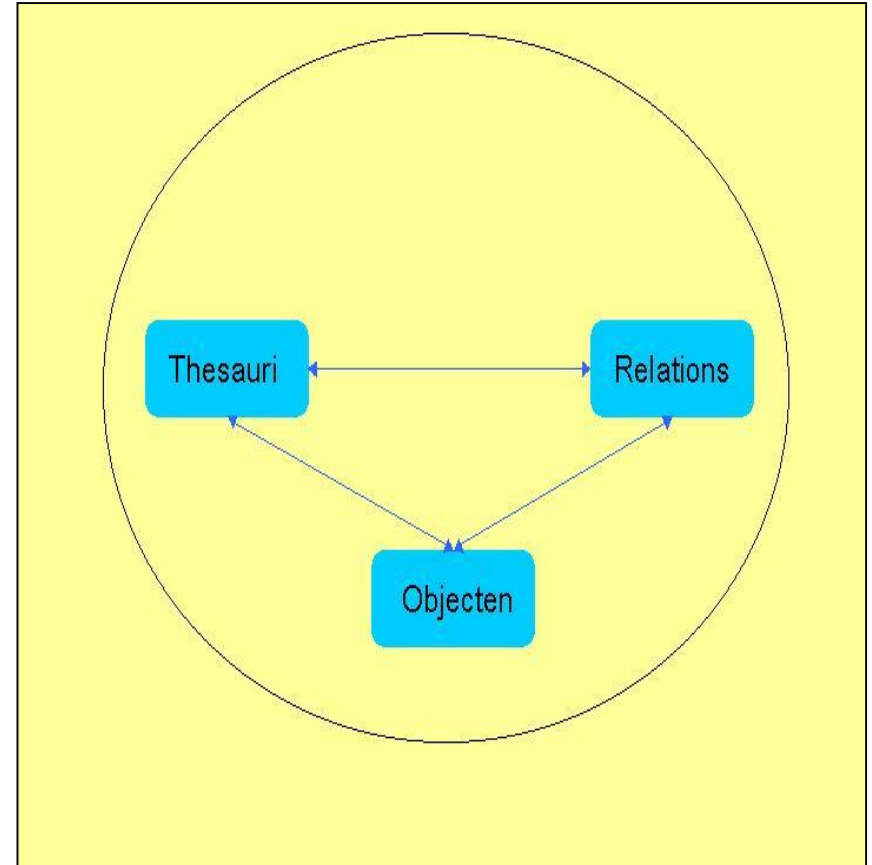


- SwiftBox (\pm 2000-2002)
- New information system to replace 6+ different systems
- NIWI – Nederlands Instituut voor Wetenschappelijk Informatievoorziening (1998-2003?)
- Merger of 6 different KNAW institutions
- Problem: conflicts between “old” metadata owners

Concepts, relations, thesauri



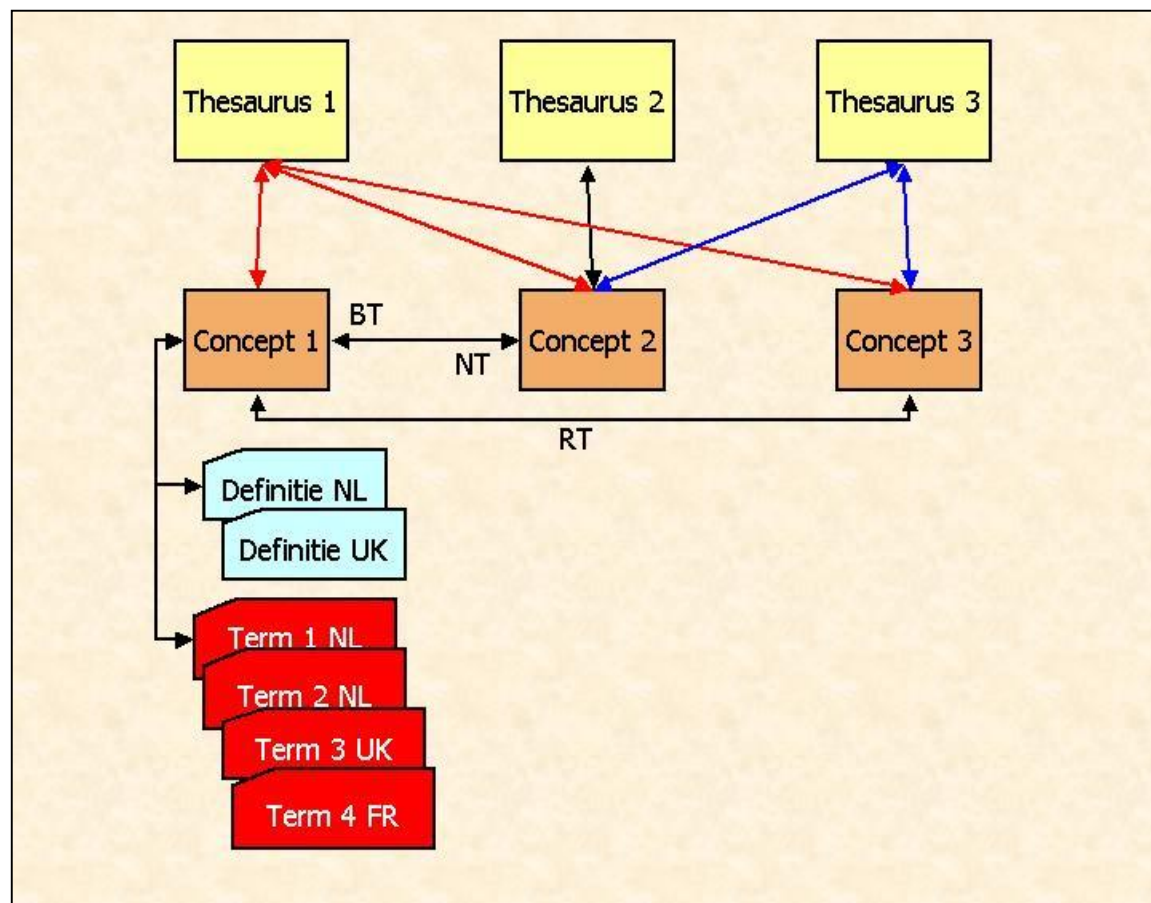
- <http://niekvanbaalen.net/swiftbox/>
- Objects
- Linked to objects
- Relations defined by Thesaurus
- Objects defined by Thesaurus



SwiftBox Thesaurus



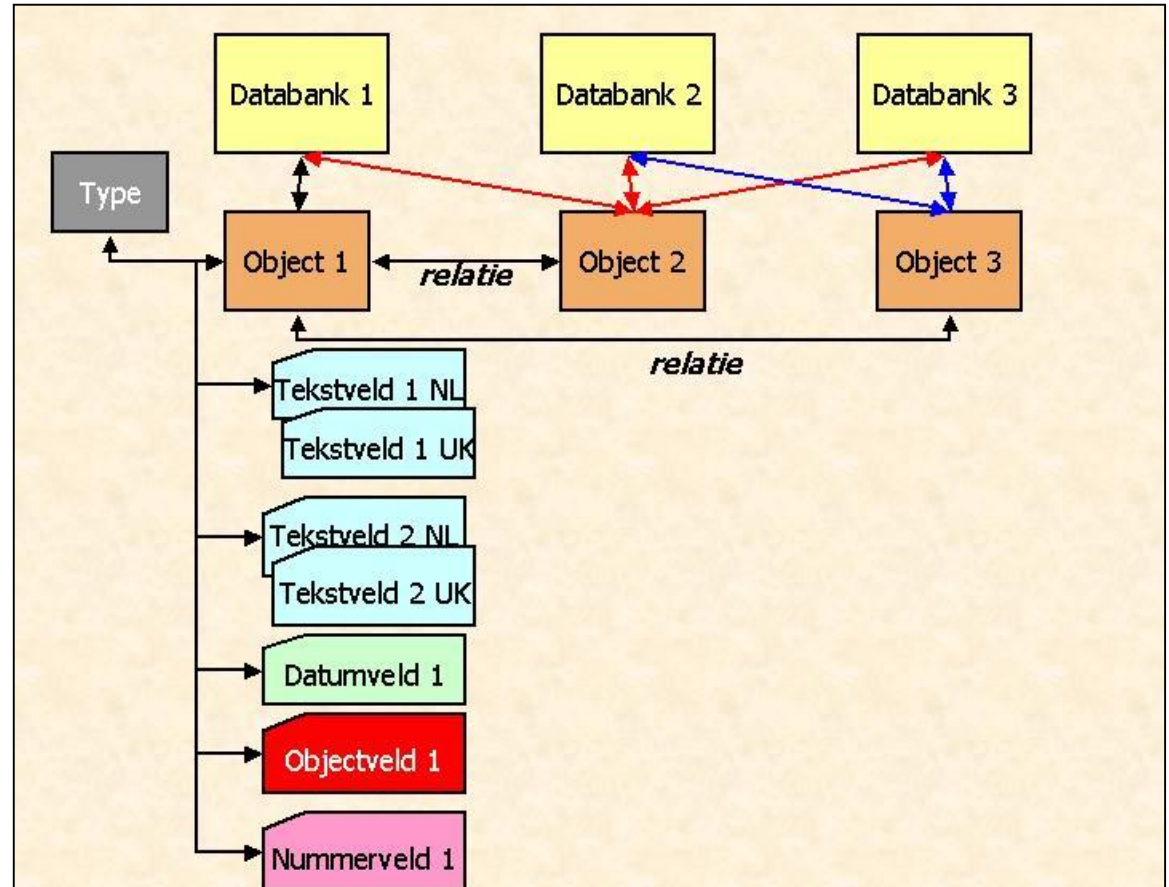
- “Abstract” concepts
- Unique identifier
- Synonyms
- Multilingual labels
- Relations:
 - RT - Related term
 - NT - Narrower term
 - BT - Broader term
- Terms in 1 or more thesauri



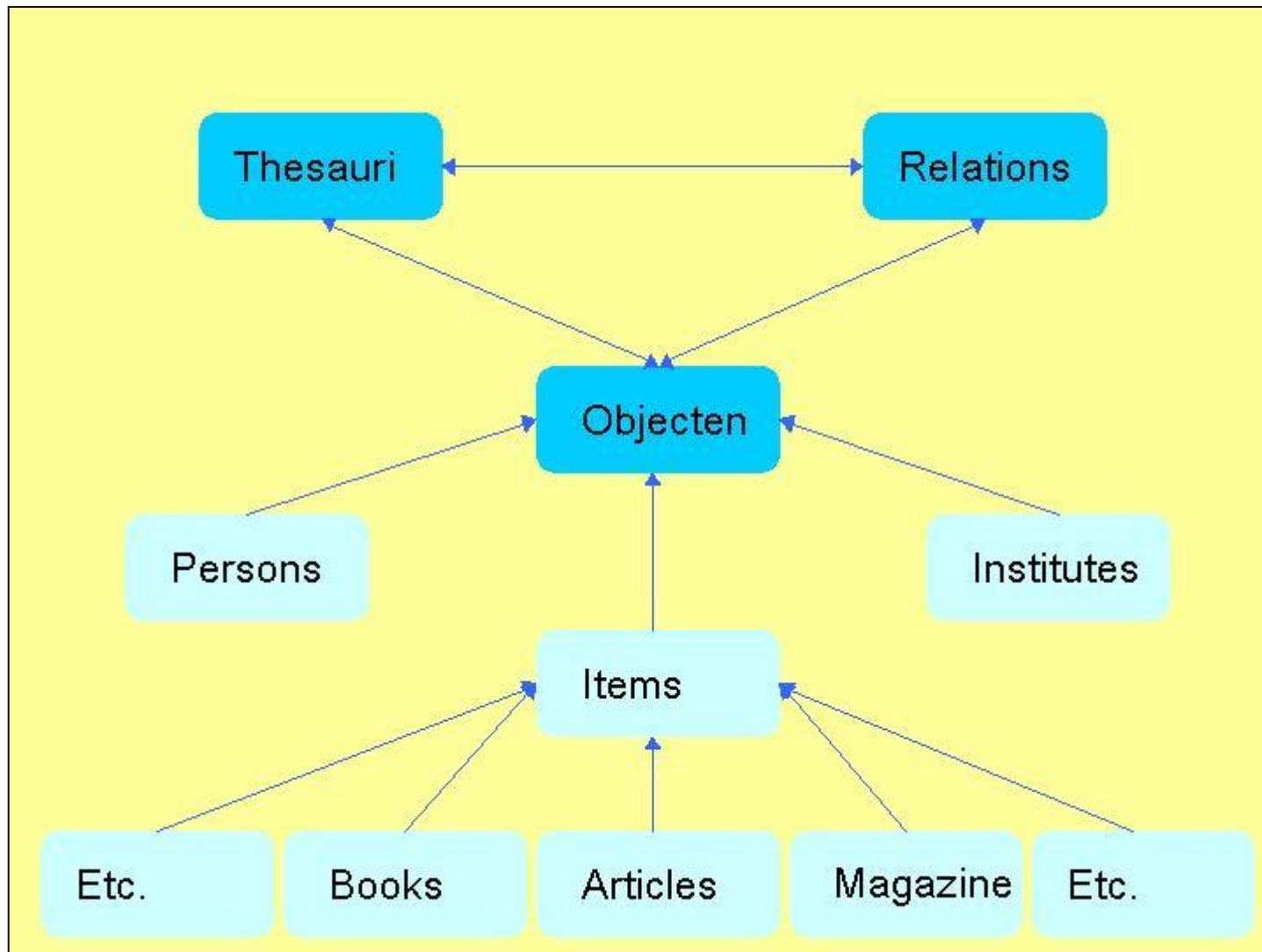
SwiftBox Object types



- Based on Thesaurus model
- Fields (attributes) are concepts defined by thesaurus term
- Fields linked to Object
- Object types: template – set of fields
- Optionally: + set of relations



SwiftBox Object types



So, what's new?



■ SwiftBox

- Objects and relations
 - Abstract concepts
 - Object types
 - Thesauri
 - Relations between databases
 - Synonyms
 - Multilingual labels
- ## ■ We even had
- Decoupled back and front end

■ Semantic Web

- Triples, RDF
- URI's
- Ontologies
- Vocabularies
- Linked Data

■ FRBR

- ERD
- ??
- Group 1 entities
 - Work
 - Expression
 - Manifestation
 - Item
- Group 2 entities
 - Person
 - Corporate body
- Group 3 entities
 - Subjects