

On Ambition and Ambiguity

Niklas Lindström & Fredrik Klingwall
National Library of Sweden

The context: Libris XL

Observations from linking The Swedish National Union Catalogue

- 14 Million bibliographic descriptions and 45 Million holdings (National bibliography and authority file)
- Libris XL is the fifth generation of Libris systems since the start 1972. (Libris is 50 years! 🎉)
- Lots of automatic pipelines for input, output and transformations.
- Since 2018 a Native Linked Data system based on BIBFRAME in an ecosystem dependent on MARC21.

Ambition

What:

- Large-scale (national & world wide) shared descriptions.
- *Maintaining* quality and consistency.

To support:

- Daily discovery (by patrons and researchers).
- Effective means of curation.
- Comprehensive and sustainable library knowledge graph.

Requires: semantic interoperability.

Within: a “fairly wide” context (the library and cultural heritage domains; the public sector...).

Inherited Problems: Abundant Conflations

We're still dealing with collapsed ambiguities extracted from MARC21.

The meaning of **bib 240** – which entities are alluded to?

If \$1 is present: the title (\$a) of the original and its language (\$1).

Else: the title of “this” work (though there is no distinct work in the bibliographic MARC21 format).

The same field alludes to two distinct things. Similar problems *throughout*. ISBN:s with qualifiers (“hbk”). Using plain text notes to “qualify” what the thing specifically is, or what a relation (link) might be). Fixed fields contradicted by additional fields.

There is **no** technical **precision of identity**, only a **richness of detail**...

Wicked Problems? 🙄

The ambition: To take us out of *thinking in* MARC21:

- Its constraints.
- Its redundant conflations.

Without:

- **keeping old conflations;**
- **creating new ambiguities.**

Normalization

In databases normalization is an **iterative process** where each step leads to more manageable information by reducing data redundancy and repetition by distributing **entities** into relational objects (tables) referenced with unique keys (Entity-Relationship Model)

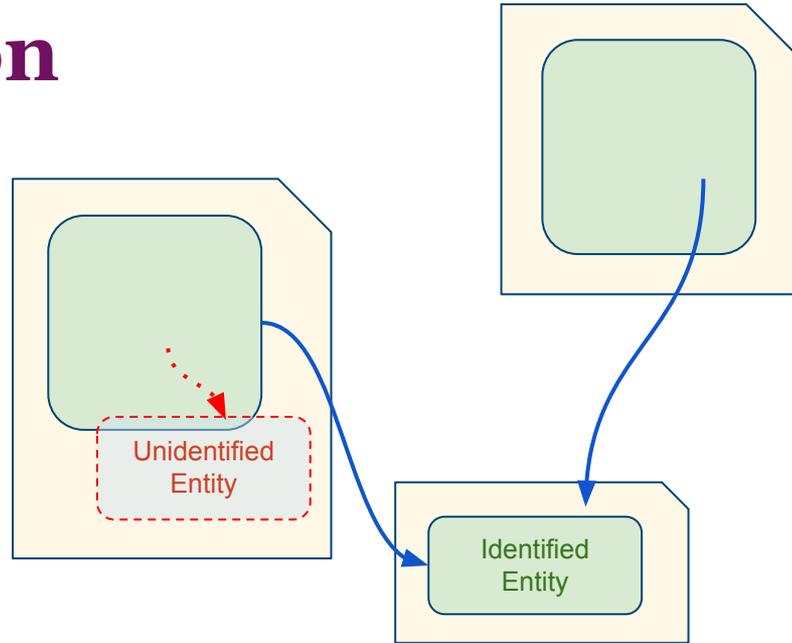
Unstructured notes and copied information is usually the opposite of normalized information.

While total normalization usually is not necessary, it provides a well-functioning information environment. The method usually accommodate new types of data without changing existing structure too much.

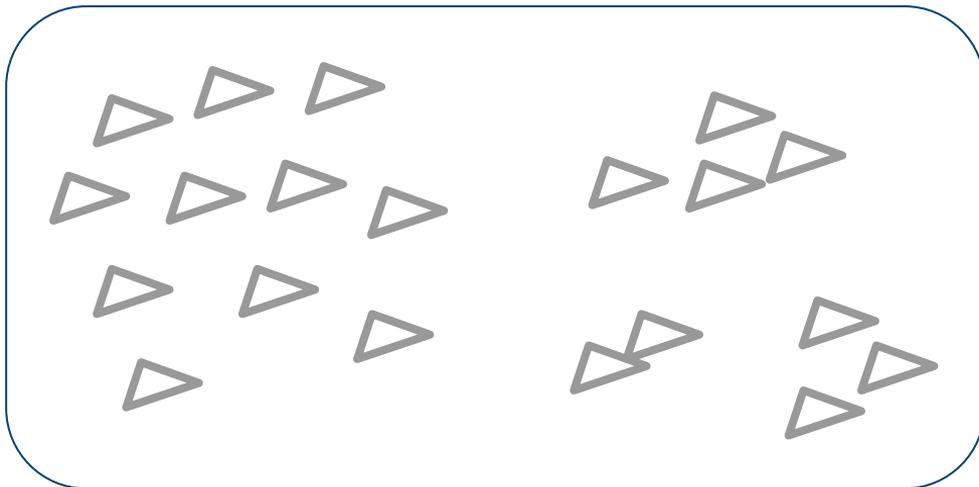
An authority file is a kind of normalization, although dependent only on string matching and in most systems features a denormalization step by copying any changes of the source back into the record.

Data Normalization

- ❑ Enables **reuse**
- ❑ Eliminates redundancy
- ❑ Ensures **consistency**

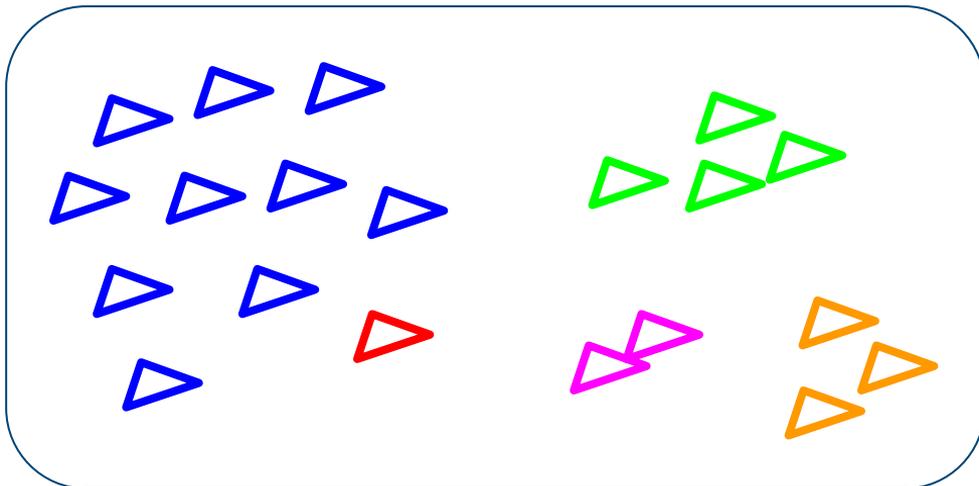


Data...



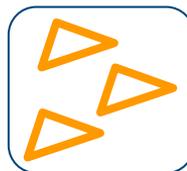
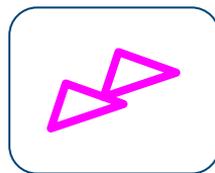
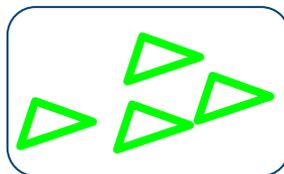
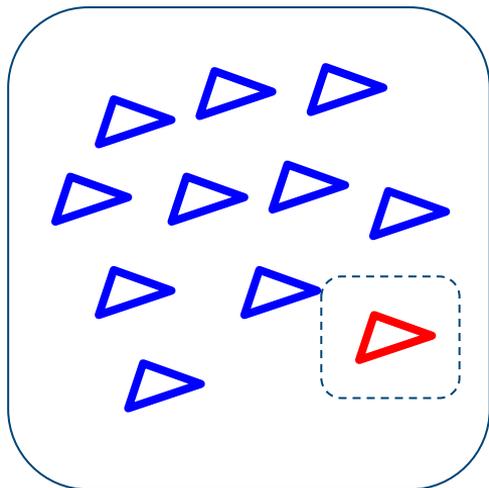
```
000 07398cam a2201729 i 4500
008 151017s2016 sw a||||j |00| 1 swe c
020 a 978-91-638-8592-1 q inbunden
041 1 a swe h eng
082 0 a 823.82 23/swe
100 1 a Carroll, Lewis, d 1832-1898 4 aut
240 1 a Alice's adventures in Wonderland. l Svenskas (Runnquist)
245 1 a Alice i Underlandet /c av Lewis Carroll ; i översättning av
Åke Runnquist ; med illustrationer av Tove Jansson
264 1 a Stockholm :b Bonnier Carlsen,c 2016
264 3 a Lettland,c 2016
300 a 122 sidor :b illustrationer (vissa i färg) ;c 25 cm
336 a text 2 rdacontent
337 a oförmedlad 2 rdamedia
338 a volym 2 rdacarrier
500 a Första svenska upplaga i denna översättning 1966
650 7 a Flickor 2 barn 0 244824
650 7 a Drömmar 2 barn 0 244767
650 7 a Parallella världar 2 barn 0 308450
650 7 a Fantasiresor 2 barn 0 250103
650 7 a Fantasiidjur 2 barn 0 248696
655 7 a Kapitelböcker 2 barngf 0 243987
655 7 a Barn- och ungdomslitteratur 2 barngf 0 352941
700 1 a Runnquist, Åke, d 1919-19914 trl 0 231599
700 1 a Jansson, Tove, d 1914-20014 ill 0 191503
```

... Statements...

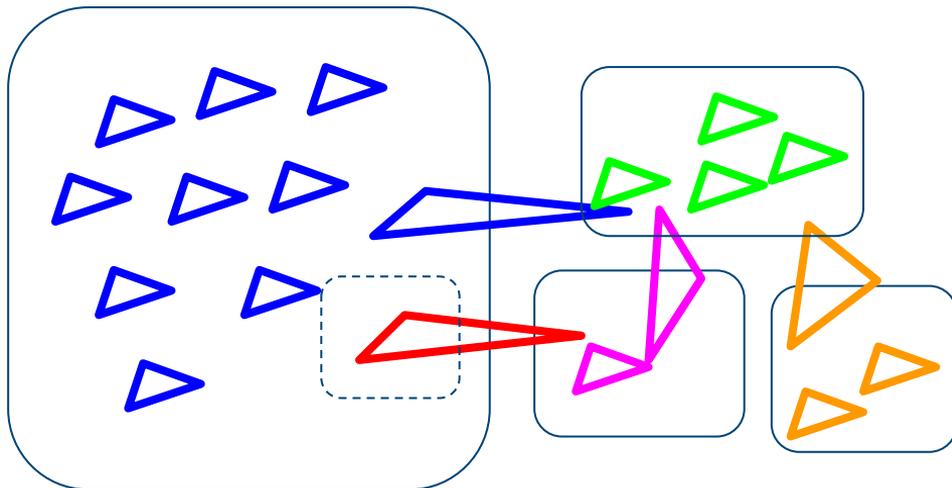


```
000 07398cam a2201729 i 4500
008 151017s2016 sw a||||j |00| 1 swe c
020 a 978-91-638-8592-1 q inbunden
041 1 a swe h eng
082 0 a 823.82 23/swe
100 1 a Carroll, Lewis, d 1832-1898 4 aut
240 1 a Alice's adventures in Wonderland. l Svenskas (Runnquist)
245 1 a Alice i Underlandet /c av Lewis Carroll ; i översättning av
Åke Runnquist ; med illustrationer av Tove Jansson
264 1 a Stockholm :b Bonnier Carlsen,c 2016
264 3 a Lettland,c 2016
300 a 122 sidor :b illustrationer (vissa i färg) ;c 25 cm
336 a text 2 rdacontent
337 a oförmedlad 2 rdamedia
338 a volym 2 rdacarrier
500 a Första svenska upplaga i denna översättning 1966
650 7 a Flickor 2 barn 0 244824
650 7 a Drömmar 2 barn 0 244767
650 7 a Parallella världar 2 barn 0 308450
650 7 a Fantasiresor 2 barn 0 250103
650 7 a Fantasidejur 2 barn 0 248696
655 7 a Kapitelböcker 2 barngf 0 243987
655 7 a Barn- och ungdomslitteratur 2 barngf 0 352941
700 1 a Runnquist, Åke, d 1919-19914 trl 0 231599
700 1 a Jansson, Tove, d 1914-20014 ill 0 191503
```

... Entities...



... Relationships!



Print extent /4nrg0ckg0vwjhl8#it

Extent

label 122 sidor

sameAs <http://libris.kb.se/resource/bib/18546404>

hasNote

- Note**

label Första svenska upplaga i denna översättning 1966
- Note**

label [Ny tryckning], 2019. 120 sidor

hasTitle

Title

mainTitle Alice i Underlandet

mediaType <https://id.kb.se/term/rda/Unmediated>

instanceOf <https://id.kb.se/term/rda/Volume>

carrierType <https://id.kb.se/term/rda/Volume>

manufacture

Manufacture

year 2016

place

Place

label Lettland

publication

PrimaryPublication

date [2016]

year 2016

agent

Agent

label Bonnier Carlsen

place

Place

label [Stockholm]

country <https://id.kb.se/country/sw>

identifiedBy

ISBN

value 9789163885921

qualifier inbunden

issuanceType [Monograph](#)

hasDimensions

Dimensions

label 25 cm

physicalDetailsNote illustrationer (vissa i färg)

responsibilityStatement av Lewis Carroll ; i översättning av Åke Runnquist ; med illustrationer av Tove Jansson

graph /q6q8c6q6n6ps91cf#lens=card

Text

subject

- <https://id.kb.se/term/barn/Flickor>
- <https://id.kb.se/term/barn/Dr%C3%B6mmar>
- <https://id.kb.se/term/barn/Parallella%20%C3%A4rldar>
- <https://id.kb.se/term/barn/Fantasiresor>
- <https://id.kb.se/term/barn/Fantasidjur>

hasTitle

Title

source </4nrg0ckg0vwjhl8#it>

mainTitle Alice i Underlandet

language <https://id.kb.se/language/swe>

genreForm

- <https://id.kb.se/marc/FictionNotFurtherSpecified>
- <https://id.kb.se/term/barn/g/Kapitelb%C3%B6cker>
- <https://id.kb.se/term/barn/g/Barn-%20och%20ungdomslitteratur>

contentType <https://id.kb.se/term/rda/Text>

contribution

source

PrimaryContribution

role <https://id.kb.se/relator/author>

agent </jyx0tr142tk2vrj#it>

- Contribution**

role <https://id.kb.se/relator/translator>

agent </fctwjbz3cspqf7#it>
- Contribution**

role <https://id.kb.se/relator/illustrator>

agent </wt79bh6f2j46dtr#it>

translationOf </ctc9w1j9lstvhbc#it>

classification

- ClassificationDdc**

code 823.8

edition full

editionEnumeration 23/swe
- Classification**

code Hcf

inScheme

ConceptScheme

code kssb

sameAs <https://id.kb.se/term/kssb/B/>

version 8

intendedAudience <https://id.kb.se/marc/Juvenile>

illustrativeContent <https://id.kb.se/marc/Illustration>

graph /ctc9w1j9lstvhbc#lens=card

Work

hasTitle

Title

mainTitle Alice's adventures in Wonderland

language <https://id.kb.se/language/eng>

marc:version (Runnquist)

graph /jyx0tr142tk2vrj#lens=card

translationOf

sameAs <http://libris.kb.se/resource/auth/44914>

rdfs:seeAlso </jgvz6wz221fmyg#it>

lifeSpan 1832-1898

givenName Lewis

familyName Carroll

description Brittsk författare, matematiklärare vid universitetet i Oxford. Pseudonym för Charles Lutwidge Dodgson.

nationality <https://id.kb.se/nationality/e-uk--->

agent

identifiedBy

- VIAF**

value 66462036
- ISNI**

value 000000012137136X

graph /fctwjbz3cspqf7#lens=card

agent on

sameAs <http://libris.kb.se/resource/auth/231599>

rdfs:seeAlso

Person

name Äqvist, Rune

lifeSpan 1919-1991

lifeSpan 1919-1991

givenName Åke

familyName Runnquist

description Förlagsman och litteraturkritiker. Skrev även under pseudonymen Rune Äqvist (eller Äqvist).

nationality <https://id.kb.se/nationality/e-sw--->

identifiedBy

- VIAF**

value 79316184
- ISNI**

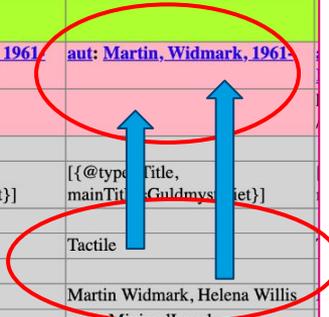
value 000000058660934

Work normalization in Libris XL

1. **Split** the bibliographic information into **bf:Instance** with a “local” **bf:Work** entity.
2. **Add precision** where possible (using heuristics to add author roles, etc.).
3. Compare works in a subset with **matching** heuristics (title, author, language, ...).
4. **Assess** remaining disparities in the resulting cluster (uncoordinated classifications, subjects, genre/form; obviously a mix of content/form details).
5. Replace the local work entities with a **link** to the resulting normalized work.

Instances of the same bf:Work?

dxq2cstq4sptgl7	j1tt17dv4dlr593	g0s845gs257f2k1	9kvzcp2n7v5gx37n	xg8922s8481bmt0	8sltrf3l39tg3kx	1g1p80rxszb0hj65
Title: guldmysteriet						
https://id.kb.se/marc/Juvenile						
Text						
https://id.kb.se/language/swe						
https://id.kb.se/marc/FictionNotFurtherSpecified	https://id.kb.se/marc/FictionNotFurtherSpecified	https://id.kb.se/marc/NotFictionNotFurtherSpecified	https://id.kb.se/marc/NotFictionNotFurtherSpecified	https://id.kb.se/marc/FictionNotFurtherSpecified	https://id.kb.se/marc/FictionNotFurtherSpecified	https://id.kb.se/marc/FictionNotFurtherSpecified
https://id.kb.se/term/barngf/Barn-%20och%20ungdomslitteratur						
https://id.kb.se/term/barngf/Deckare						
https://id.kb.se/term/barngf/Kapitelb%C3%B6cker						
https://id.kb.se/term/barn/Banker						
https://id.kb.se/term/barn/Guld						
https://id.kb.se/term/barn/Tjuvar						
91	91	91	91	96	96	160
839.738 kssb, 8: Hcf	839.738 kssb, 8: Hcf					
aut: Martin, Widmark, 1961-ill: Helena, Willis, 1964-	aut: Martin, Widmark, 1961-ill: Helena, Willis, 1964-	aut: Martin, Widmark, 1961-ill: Helena, Willis, 1964-	aut: Martin, Widmark, 1961-ill: Helena, Willis, 1964-	aut: Martin, Widmark, 1961-ill: Helena, Willis, 1964-	aut: Martin, Widmark, 1961-ill: Helena, Willis, 1964-	aut: Martin, Widmark, 1961-ill: Helena, Willis, 1964-
https://id.kb.se/term/rda/Text	https://id.kb.se/term/rda/Text					
{{@type=Title, mainTitle=Guldmysteriet}}						
Print	Instance	Electronic	Electronic	Electronic	Electronic	Tactile
1. uppl. i finare färg	1. uppl.					
Martin Widmark, Helena Willis	Martin Widmark, Helena Willis					Martin Widmark, Helena Willis
marc:FullLevel	marc:FullLevel	marc:AbbreviatedLevel	marc:AbbreviatedLevel	marc:AbbreviatedLevel	marc:AbbreviatedLevel	marc:MinimalLevel
2016 • Bonnier Carlsen • Stockholm • https://id.kb.se/country/sw	2006 • Bonnier Carlsen • Stockholm • https://id.kb.se/country/sw	2018 • Bonnier Carlsen • https://id.kb.se/country/sw	2019 • Bonnier Carlsen • https://id.kb.se/country/sw	2014 • Bonnier Carlsen • https://id.kb.se/country/sw	2015 • Bonnier Carlsen • https://id.kb.se/country/sw	2006 • TPB • Enskede • https://id.kb.se/country/sw
9789163889455 • inbunden	9163848724 • inb. 9789163848728 • inb.	9789178031597 • PDF	9789178033904 • EPUB	9789143511505	9789163883132	
91 s.	91, [2] s.	91 sidor	91 sidor	96 s.	96 s.	2 vol. (160 s. tryckt punktskrift)



The Ongoing Challenge of Ambiguity

When presented with **ambiguous information** a human can often, within certain contexts and limits, still easily determine "what's going on". If we want to share descriptions though, *ambiguities become a hard problem*.

(such as indistinct identities and omitted or implicitly indirect relations)

Unless you refer to certain **distinct, identifiable entities**, the meaning is diffused.

(such as the original text and its conventional genre, a translation by a specific author, the form of a printed edition, or the series of all newspaper issues)

You need not be specific, but you need to be **precise**.

(such as explicitly speaking of the *common idea* of a story, or the *idea of the translation*, rather than its *various forms* of expression or manifestations/instantiations).

The Risk of Abstraction Collapse

Ambiguity is not conflation. It is, however, **uncertainty**. And therein lies **potential conflation** of things.

A sender may intend something general, or abstract, but if that is not explicit, the receiver may interpret that as something specific, and make wrong inferences from that.

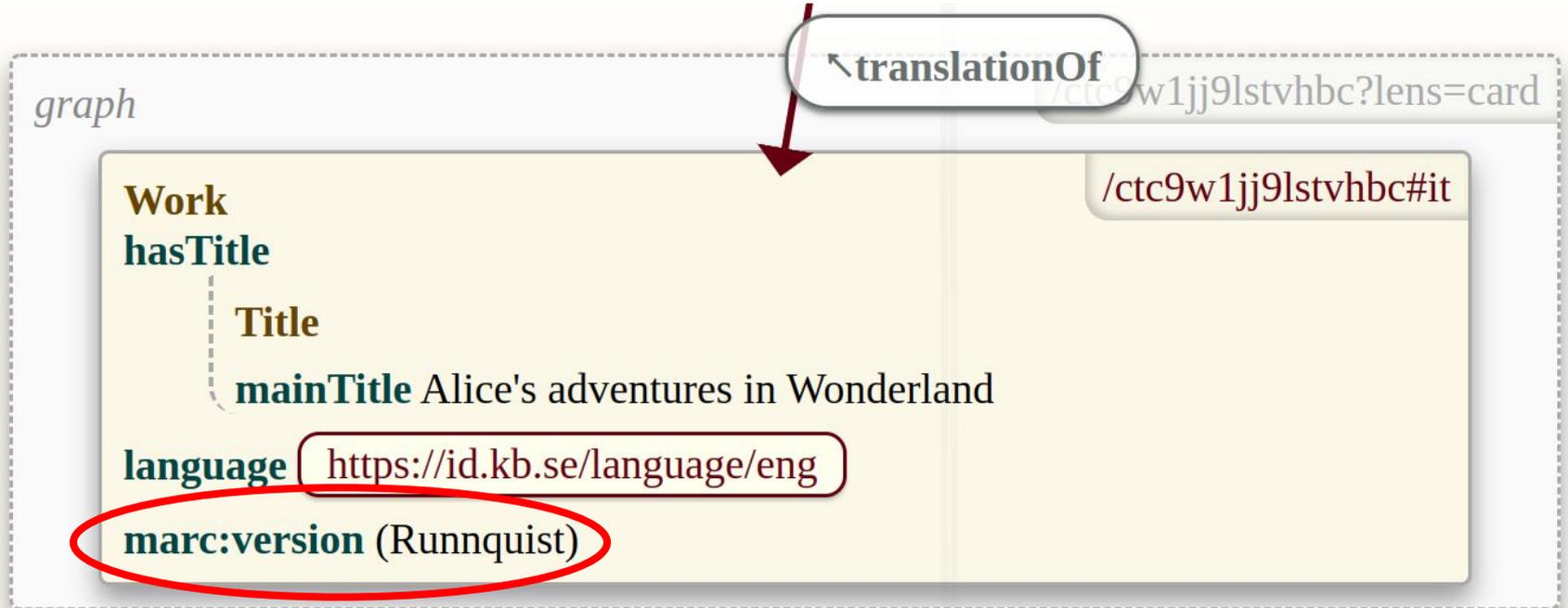
Ambiguity *may* be good (it is actually essential for communication, because humans deal with patterns, not atoms).

Machines can also deal with ambiguity (c.f. Machine Learning, e.g. entity recognition), but require more raw data to be able to disambiguate through recurring patterns.

If an ambiguity “collapse” in usage, it will lead to conflation.

Both humans and machines get into trouble when conflations happen. When being on the wrong level of detail, wrong facts are inferred, and consequently, wrong choices are made.

Collapse Into Conflation



Catering for a display detail (bib 240 \$s), at the expense of a legible, sufficiently general and reusable entity...

Precision Of Identity

In order to *efficiently* leverage RDF, and BIBFRAME upon that, as a means for interoperable fact keeping, the **ambition** needs to be higher in terms of ***precision of identity, rather than richness of detail.***

In MARC, those ambitions appears identical; but once you can concretize and link entities, this assumption does not hold. Not all details apply to the same thing.

Not all details are even disambiguating enough to be usable. Some of it is *just detrimental noise.*

Which are *the differences that make a difference?*

Avoid the *vagueness of theoretical abstraction* by seeking **concrete generalizations.**

Converge on the **recognisable generality of something.** Defer the theoretical definition of that until that is concrete enough to be **usable.**

WYSIWYC

What You See Is What You **Comprehend**

(Really. We do *not* comprehend what we see.)

The trick then, is to make them equal. The *message* (the data) and the *interpretation* (the usage thereof).

You know an entity by its expected details and relations.

Perception and comprehension correlate through **expectation**.

NB PRINTED, INSTANCE • MONOGRAPH, UNMEDIATED, VOLUME

Alice i Underlandet

Identifier ISBN 9789163885921 • Inbunden

Instance of [Alice i Underlandet • Swedish](#)

Creative responsib... q6qn8cq6n6ps91cf

Publication

Manufacture

Show more (3)

INSTANCE ADMIN

TEXT, WORK

Alice i Underlandet

Translation of [Alice's adventures in Wonderland • English • \(Runnquist\)](#)

Contributor ... Carroll Lewis 1832-1898 • Author, ...

Language inf... [Swedish](#)

Genre/form [Fiction \(not further specified\) • marc Kapitelböcker • barnngf ...](#)

Classification 823.8 • full • 23/swe, Hcf • kssb

TEXT, WORK

Alice i Underlandet

Translation of [Alice's adventures in Wonderland • English • \(Runnquist\)](#)

Contributor and r... ctc9wt1j9lsthvbc 001 • Illustrator

Language informa... [Alice's adventures in Wonderland](#)

Genre/form Marc:version (Runnquist)

Classification

Subject Language inf... [English](#)

Intended audience [Juvenile • marc](#)

Content type [text](#)

Illustrative conten... [illustrations](#)

How to Maintain Comprehension?

How do we maintain *equal expectations* of – and *shared identities* in – our descriptions, across roles, organizations or communities?

If developers, designers and patrons do not **share useful notions**, how can we achieve wider interoperability?

Through...

- ... Standards?
- ... Diligent maintenance, error correction and effective feedback loops?
- ... Indirected transformation chains and user-facing API:s?
- ... Application Profiles? Validation? Shapes?
- ... Concrete usage in actual applications? (Selected details in cards and chips.)

Practical Semantic Interoperability

- Act local (needs), think global (standards).

What is local, what is global? Just core RDF structures? Do you link across your local systems? Do they share terminology? Do you compare ontologies?

- Improve ontological quality (not just “old terms in new clothes”); useful concretizations; consistency & constraints.

In actual application! Metadata and user experience are two sides of the same coin.

Keep assessing if our standards meets our needs

`type: Text;`
`content: Text;`
`genreForm: Fiction.`

`type: Print;`
`issuance: Monograph;`
`media: Unmediated;`
`carrier: Volume.`

So... a printed book of some text?

It adds *richness of detail*. But does it give us more *precision of identity*?

Is it **a Hardback of a Novel** maybe?

Thank you!

niklas.lindstrom@kb.se | fredrik.klingwall@kb.se

