

# Conversion

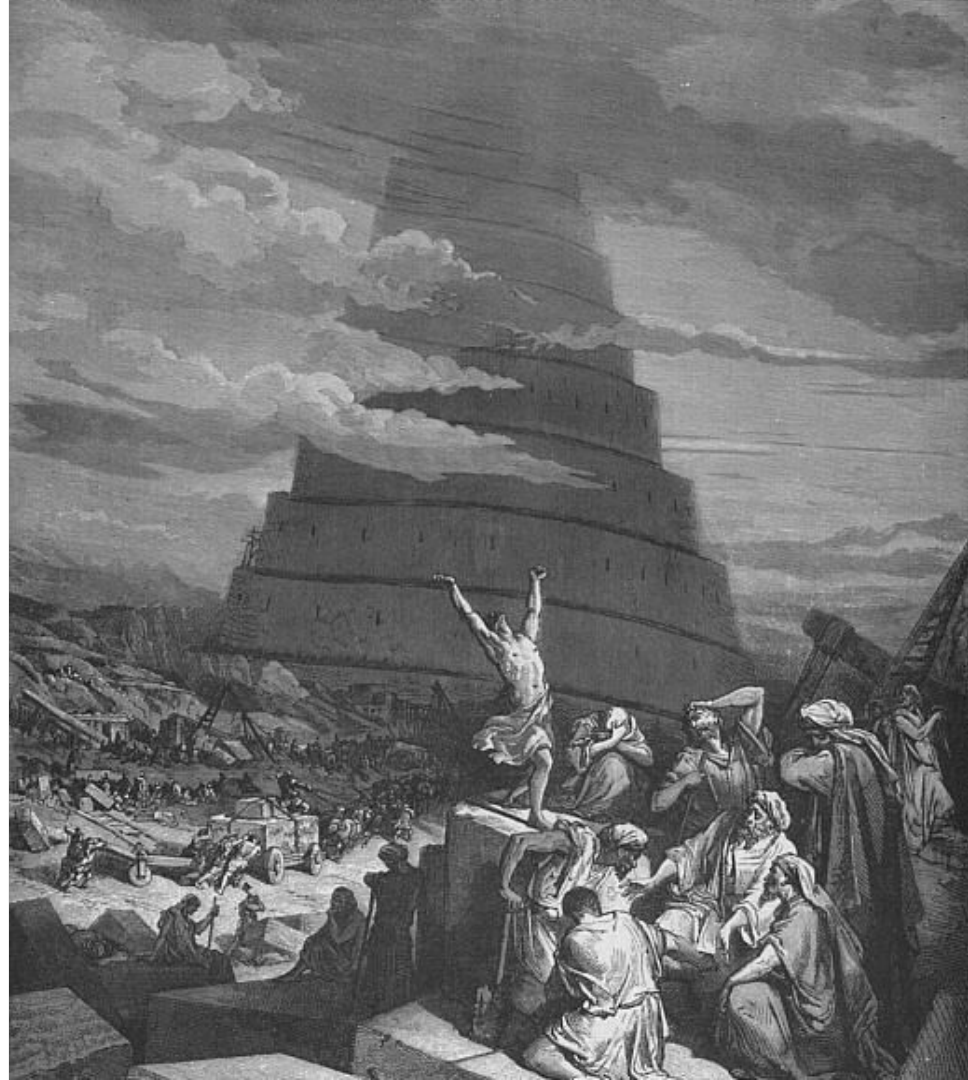
Experience From Libris, The Swedish Union Catalogue

Niklas Lindström



# WHY?

(not) MARC? RDF?



# MARC vs RDF?

*The decision we are facing in terms of bibliographic data is often couched in terms of "MARC vs. RDF", however, **that is not the actual question** that underlies that decision.*

*Instead, the question should be couched as: **entities and relations, or not?** if you want to **share entities like works and persons**, and if you want to create **actual relationships** between bibliographic entities, something other than MARC21 is required.*

– Karen Coyle <<http://kcoyle.blogspot.se/2017/05/two-frbrs-many-relationships.html>>

# Two Perspectives: Logical and Technical

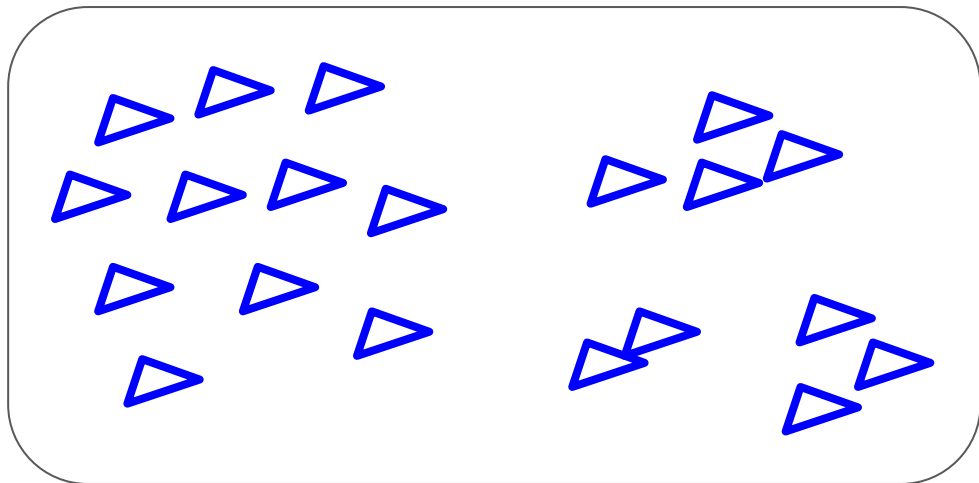
Everything is **RDF**.

Interlinked entities with structured (bnode) or literal property values.

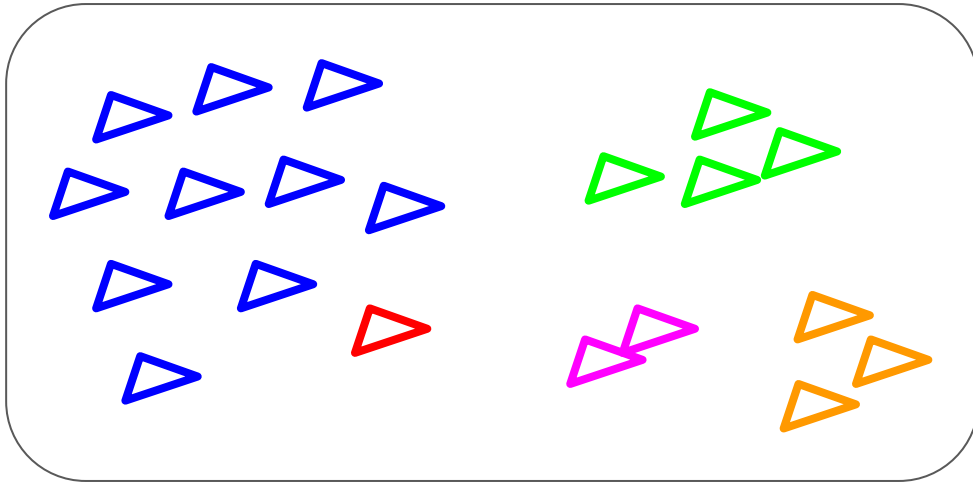
*The Vocabulary is the rule set:*

- Properties and Types
- Domains and Ranges
- Super- and Sub-

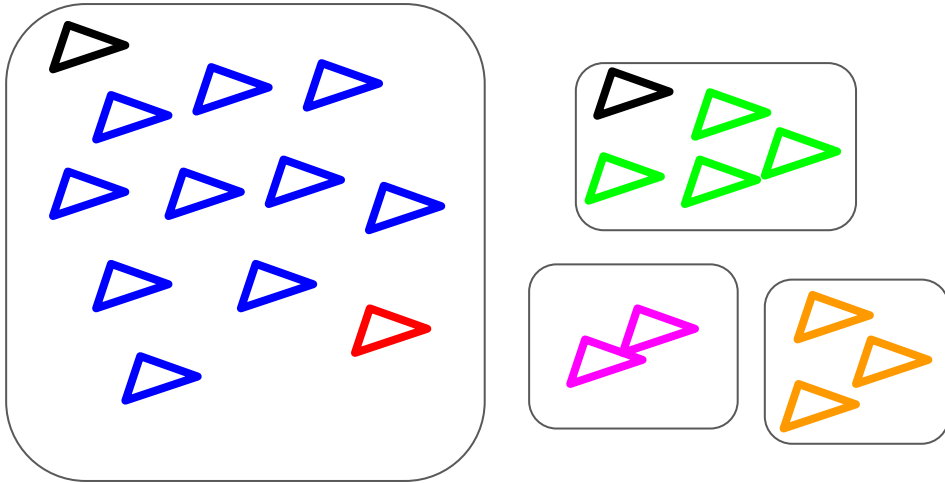
# Data



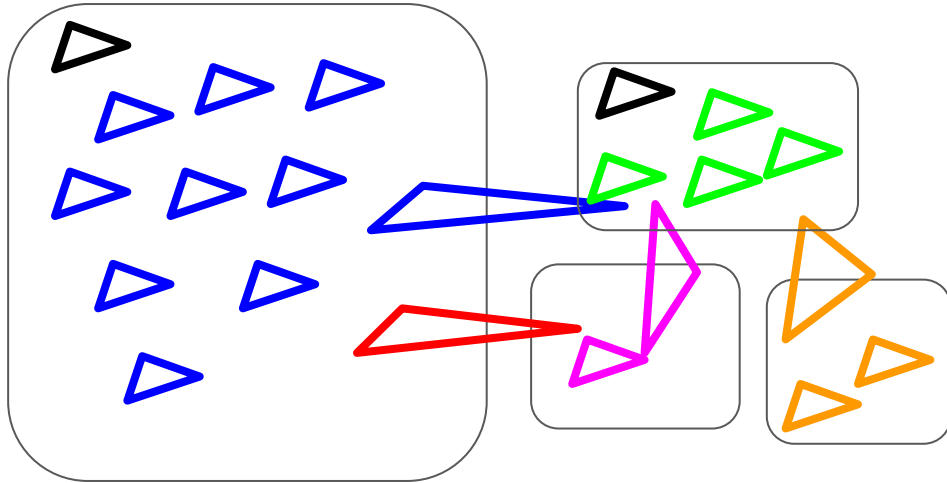
# Statements (subjects with values of predicates)



In “records” (administrative description units)



# Some Statements are Links





# Row Upon Row of Facts

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

# A Distinct Entity: the instance

- 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. I  
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 barn 0 243987
- 655 7 a Barn- och ungdomslitteratur 2 barn 0  
352941
- 700 1 a Runnquist, Åke, d 1919-19914 trl 0 231599
- 700 1 a Jansson, Tove, d 1914-20014 ill 0 191503

# The Instance Described

```
</inst/aiu> a :Instance;  
  :instanceOf </text/aiu>;  
  :issuance :Monograph;  
  :identifiedBy [ a :Isbn; :value "9789163885921"; :note "inbunden" ];  
  :title [ a :TitlePage; :mainTitle "Alice i Underlandet" ]; :responsibilityStatement "av Lewis Carroll; i  
översättning av Åke Runnquist; med illustrationer av Tove Jansson";  
  :publication [ :agent </org/bonnier-carlsen>; :date "2016"; :place </ort/stockholm> ];  
  :manufacture [ :date "2016"; :place </land/lv> ];  
  :extent [ :label "122 sidor" ];  
  :otherPhysicalDetails "illustrationer (vissa i färg)";  
  :media </rda/media/unmediated>;  
  :carrier </rda/carrier/volume>;  
  :note [ :label "Första svenska upplaga i denna översättning 1966" ];  
  :dimensions [ :label "25 cm" ];  
  :publicationCountry </country/sw> .
```

# The Work

```
</text/aiu> a :Text;  
  :title [ a :Title; :mainTitle "Alice i Underlandet" ];  
  :expressionOf </text/aaiw> .  
  :contribution [ a :PrimaryContribution; :agent </person/lwscrll6421>; :role </loc/relator/aut> ],  
    [ a :Contribution; :agent </person/krnngst56779>; :role </loc/relator/trl> ],  
    [ a :Contribution; :agent </person/tvjnssn94724>; :role </loc/relator/ill> ];  
  :content </rda/content/text>;  
  :genreForm </enum/FictionNotFurtherSpecified>,  
    </term/barngf/Barn-+och+ungdomslitteratur>, </term/barngf/Kapitelböcker>;  
  :classification </dewey/class/e23/823.8>;  
  :illustrativeContent </enum/Illustration>;  
  :intendedAudience </enum/Juvenile>;  
  :language </language/swe>;  
  :subject </term/barn/Drömmar>, </term/barn/Fantasidjur>, </term/barn/Fantasiresor>,  
</term/barn/Flickor>, </term/barn/Parallella+världar> .
```

# Related Entities

```
</text/aaiw> a:Text;  
  :title [ a :Title; :mainTitle "Alice's adventures in  
Wonderland." ];  
  :contribution [ a :PrimaryContribution;  
    :agent </person/lwscrll6421>;  
    :role </loc/relator/aut> ];  
  :language </language/eng> .
```

```
</person/lwscrll6421> a :Person;  
  :familyName "Carroll";  
  :givenName "Lewis";  
  :lifeSpan "1832-1898" .
```

```
</person/tvjnssn94724> a :Person;  
  :familyName "Jansson";  
  :givenName "Tove";  
  :lifeSpan "1914-2001" .
```

```
</term/barn/Drömmar> a :Topic ;  
  :inScheme </term/barn> ;  
  :prefLabel "Drömmar" .
```

```
</term/barngf/Barn-+och+ungdomslitteratur> a  
:GenreForm ;  
  :inScheme </term/barngf> ;  
  :prefLabel "Barn- och ungdomslitteratur" .
```

**bib**

**auth**

*events*

*vocab!*

*works*

*agents*

*roles*

*instances*

*topics*

*materials*

*genres*

*forms*

**hold**

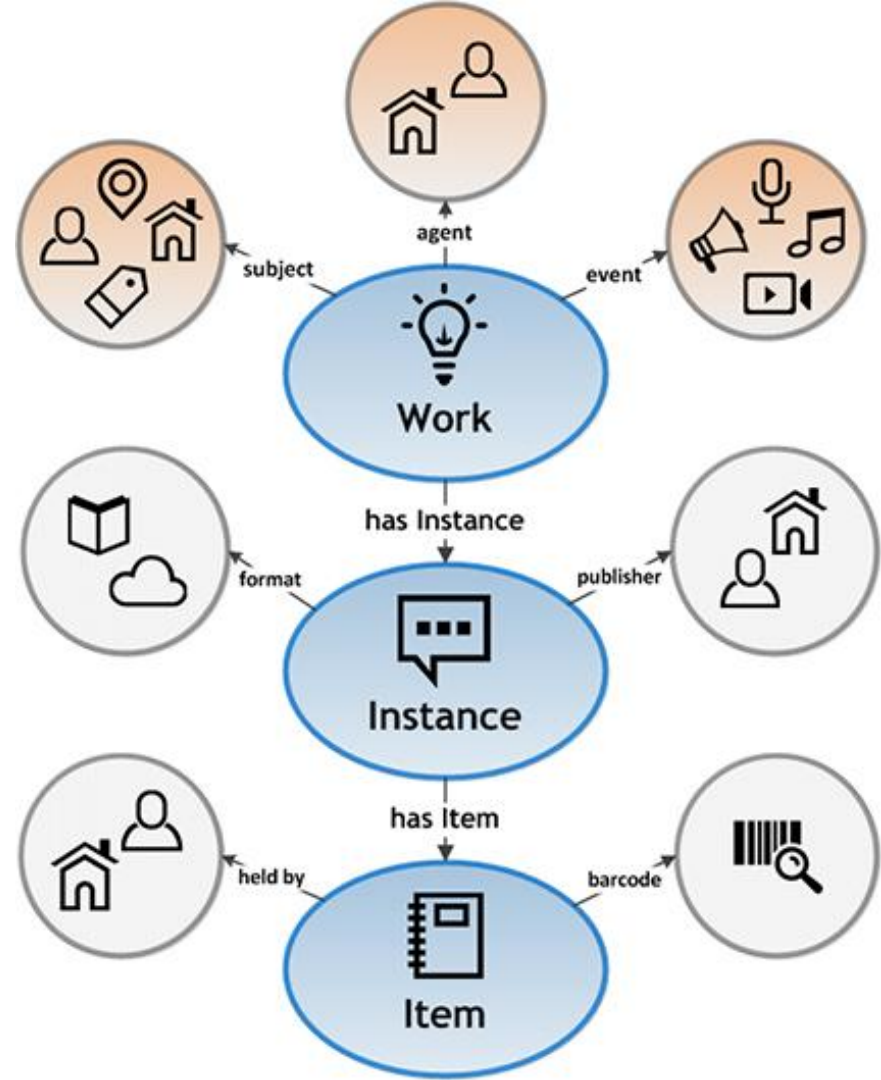
*items*

# BibFrame 2

<<http://www.loc.gov/bibframe/>>

Library of Congress  
RDF-based format.

Created to replace  
MARC21.



# Libris XL: Not Just BibFrame 2

We defined: **KBV**. Our own application ontology, for our immediate needs.

Mainly BF2 equivalencies at the core. (Plus some BFLC, RDA, MADS...)

We did the alignment midway.

The LoC mappings (both docs and XSLT transforms) were informative for us in understanding “what goes where”. We’ve needed to deviate here and there, and certain “why’s” remain to be determined.



# More Structure

Abstract base classes (e.g. Endeavour, Identity, StructuredValue, QualifiedRole).

Explicit domains/ranges/restrictions where applicable;  
domainIncludes/rangeIncludes when more loose guides are needed.

(Companion Display Declarations: Lists of interesting predicates by type (based on FRESNEL). Three main lenses: “chips”, “cards” and “page”.)

# Added Details

Added details as needed, such as:

PrimaryPublication (bib 008).

SeriesMembership (qualified form of inSeries).

# Two Perspectives: Logical and Technical

Everything is **JSON-LD** internally.

Used as “just JSON” in code (Java/Groovy, JS and Python). Stored in **Postgres**. Indexed in **ElasticSearch**. Edited using **Vue.js**.

Application code use JSON directly, guided by the ***semantics***.

All public features (editing, searching, displaying) *comply with the logical perspective*.

# Digesting MARC

Digging through a legacy of disparate  
details

“bib” turns into data about instances  
and works... ;

“hold” into items (and  
“compounds”...);

“auth” is both extended and  
dissolved into reusable terms,  
enumerations and “all other”  
entities...

... including works.

# Disseminating MARC

Coping with the inertia of local  
systems

Our implementation (*marcframe*) is driven by a declarative spec of mappings.

*It works in both directions.*

The implementation is somewhat complex...

# MarcFrame Conversion Process

1. Create entities from the MARC-data. Yields a rather deep tree of items, with some precomputed @id:s.
2. Post-process to consolidate and merge some outstanding duplicates.
3. Find linked entities (initially using out-of-band authority @id:s coupled with the record in Voyager).
4. Canonicalize @id:s by using sameAs information in our system.

# From MARC to LD and back again

```
[
  {264: {ind1: , ind2: 1,
        subfields:[
          {a: Stockholm :},
          {b: Bonnier,},
          {c: 1996}
        ]}},
  {264: {ind1: , ind2: 3,
        subfields:[
          {a: Finland}
        ]}}
]
```

```
publication: [
  {@type: Publication,
   place: {@type: Place,
          label: Stockholm},
   agent: {@type: Agent,
          label: Bonnier},
   date: 1996}
],
manufacture: [
  {@type: Manufacture,
   place: {@type: Place,
          label: Finland}}
]
```

# Combining Parts (bib 008 and 260/264)

```
{
  source: {
    marc:primaryProvisionActivity: {
      @type: PrimaryProvisionActivity,
      marc:publicationStatus: marc:SingleKnownDateProbableDate,
      year: 1977
    },
    publication: [ {@type: Publication, year: 1977} ]
  },
  result: {
    publication: [ {@type: PrimaryPublication, year: 1977} ]
  }
}
```



# MarcFrame Excerpt: bib 772

```
772: {
  aboutEntity: ?work,
  addLink: supplementTo,
  resourceType: Work,
  pendingResources: {
    _:work: {link: expressionOf, resourceType: Work},
    _:instanceOf: {link: instanceOf, resourceType: Work},
    _:hasInstance: {link: hasInstance, resourceType: Instance},
    _:contribution: {addLink: contribution, resourceType: PrimaryContribution, embedded: true},
    _:agent: {about: _:contribution, addLink: agent, resourceType: Agent},
    _:title: {addLink: hasTitle, resourceType: Title, embedded: true},
    _:hasInstanceTitle: {about: _:hasInstance, addLink: hasTitle, resourceType: Title, embedded: true},
    _:workTitle: {about: _:work, addLink: hasTitle, resourceType: Title, embedded: true},
    _:instanceOfContribution: {about: _:instanceOf, addLink: contribution, resourceType: PrimaryContribution,
embedded: true},
    _:instanceTitle: {about: _:instanceOf, addLink: hasTitle, resourceType: Title},
    _:provisionActivity: {about: _:hasInstance, link: provisionActivity, resourceType: ProvisionActivity,
embedded: true},
    _:hasInstanceDescribedBy: {about: _:hasInstance, addLink: describedBy, resourceType: Record},
    _:describedBy: {addLink: describedBy, resourceType: Record}
  },
},
```

```
il: {
  property: marc:toDisplayNote,
  tokenMap: {0: true, 1: false},
  marcDefault: 0
},
$a: {about: _:agent, property: label},
$b: {about: _:hasInstance, property: editionStatement},
$c: {about: _:title, property: qualifier},
$d: {about: _:hasInstance, property: provisionActivityStatement},
$g: {about: _:hasInstance, addProperty: part},
$h: {about: _:hasInstance, addLink: extent, resourceType: Extent, property: label},
NOTE:?$i: {addLink: relationship, resourceType: Relationship, property: label},
$i: {addProperty: marc:displayText},
$k: {about: _:hasInstance, addProperty: seriesStatement},
$m: {about: _:hasInstance, property: marc:materialSpecificDetails},
$n: {about: _:hasInstance, addLink: hasNote, resourceType: Note, property: label},
$r: {about: _:hasInstance, addLink: identifiedBy, resourceType: ReportNumber, property: value},
$s: {about: _:workTitle, property: mainTitle},
$t: {about: _:hasInstanceTitle, property: mainTitle},
$u: {about: _:hasInstance, addLink: identifiedBy, resourceType: STRN, property: value},
$x: {
  NOTE:targetMatch: 022,
  about: _:hasInstance, addLink: identifiedBy, resourceType: ISSN, property: value
},
$y: {about: _:hasInstance, addLink: identifiedBy, resourceType: CODEN, property: value},
$z: {
  NOTE:targetMatch: 020,
  about: _:hasInstance, addLink: identifiedBy, resourceType: ISBN, property: value
},
$w: {about: _:hasInstanceDescribedBy, property: controlNumber, uriTemplate: http://libris.kb.se/bib/{\_}, matchUriToken: ^\\d{1,14}$
},
$6: {property: marc:fieldref},
$7: {property: marc:controlSubfield},
$8: {property: marc:groupid},
$9: {addProperty: partNumber},
subfieldOrder: 6 7 i a c t b e d h s k x z w 4 8 9
```

# Why Is It Hard To Exit MARC21?



# “What Catalogers Like/Love About MARC”

- MARC is Familiar
- ... is WYSIWYG
- ... Gives Catalogers Control
- ... is International
- ... is the Lingua Franca of Cataloging
- **... is Concise**
- **... is Very Detailed**
- **... is Flat**
- ... is Extensible
- ... has Mnemonics
- **... is Finite and Authoritative**

– Karen Coyle <<http://kcoyle.blogspot.se/2017/04/if-it-aint-broke.html>>

# (Too?) High Granularity (for Repeatability)

```
020 $a 91-0-056322-6 $q (inb.)
```

```
isbn: "91-0-056322-6"
```

```
identifiedBy: [  
  {  
    @type: "ISBN",  
    value: "91-0-056322-6",  
    qualifier: "inb.",  
  }  
]
```

# Records Within Records (Within Records)

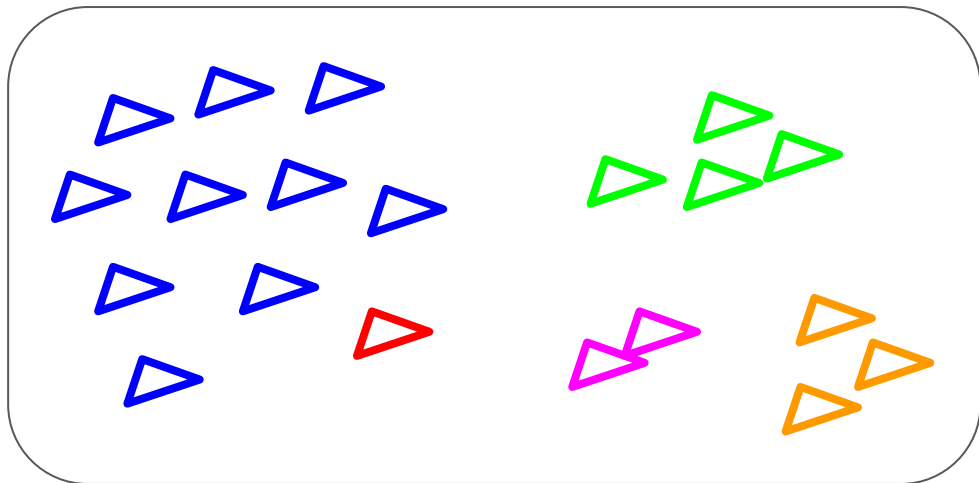
Bib 765 (original language entry):

a Main entry heading (NR) 1XX of Link Target  
b Edition (NR) 250 of Link Target  
c Qualifying information (NR)  
d Place, publisher, and date of publication (NR) 260 #c of Link Target  
g Related parts (R) (se also #t). See also 760#i  
k Series data for related item (R) 4XX or 8XX  
m Material-specific details (NR)  
n Note (R)  
o Other item identifier (R)  
r Report number (R) 088  
s Uniform title (NR) 240  
t Title (NR) 222 or 245  
u ISRN or other standard technical report number (NR) 027

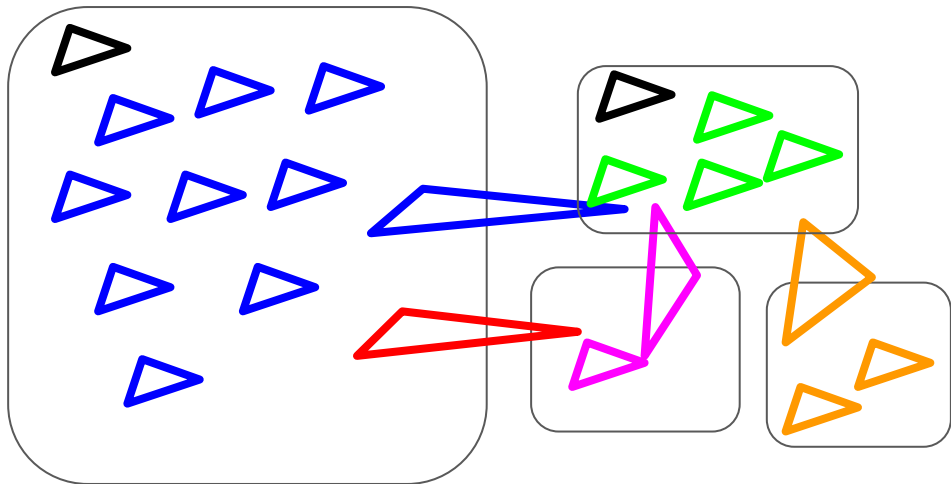
# Hidden Complexity: Tangled Entities

```
{772: {ind1: 0, ind2: ,  
  subfields: [  
    {s: Main Thing},  
    {z: 00-0-000000-0}  
  ]  
}}
```

```
instanceOf: {  
  @type: Text,  
  supplementTo: [{  
    @type: Work,  
    expressionOf: {  
      @type: Work,  
      hasTitle: [{@type: Title,  
        mainTitle: Main Thing}]  
    },  
    hasInstance: {  
      @type: Instance,  
      identifiedBy: [{@type: ISBN,  
        value: 00-0-000000-0}]  
    }  
  ]  
}
```







# Overcoming Barriers



# KBV + “Plan B”

Our type hierarchy is less clean (closer to bib 000, 007 and 008).

We keep the remainder using separate, MARC-like properties and classes.

Some pieces are necessary but poorly understood.

The rest is just kept to be *cleaned up or eventually discarded*.



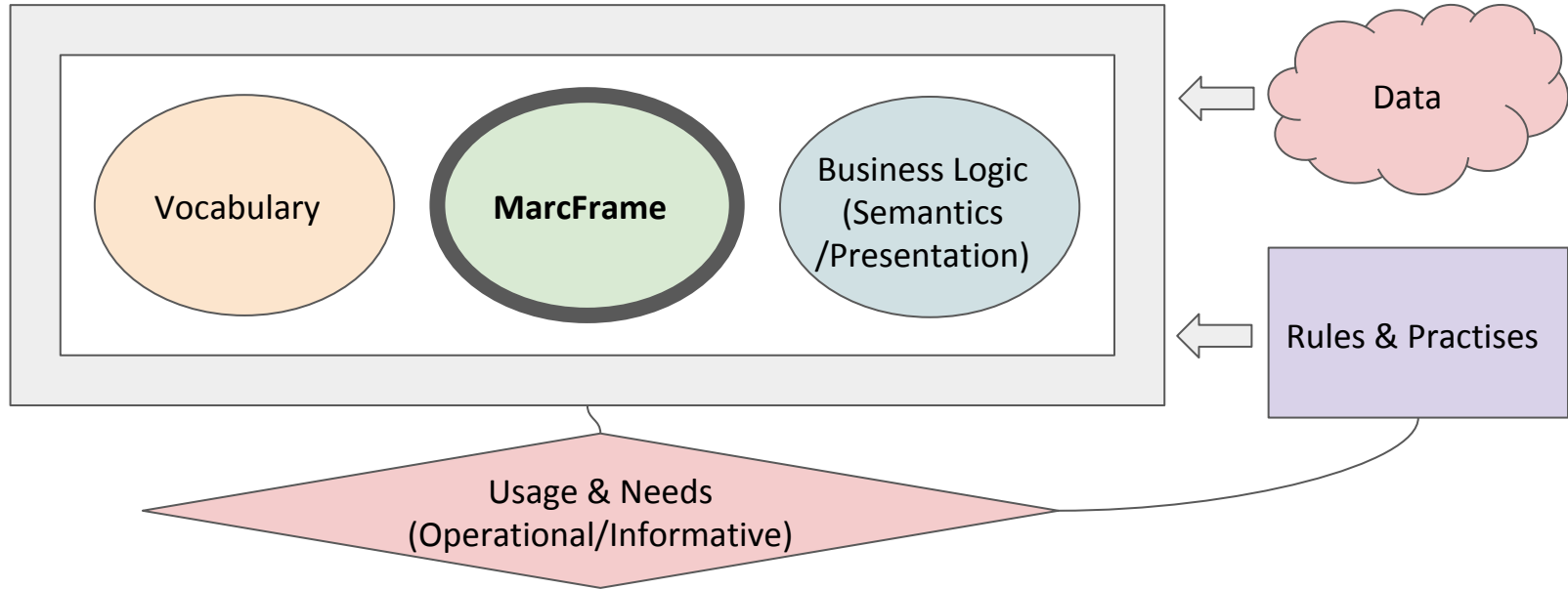
DÈS JUIN,  
6 NOUVEAUX SETS  
À DÉCOUVRIR !

LEGO JURASSIC WORLD™

31

Jurassic World is a trademark and copyright of Universal Studios and  
Amblin Entertainment, Inc. Licensed by Universal Studios Licensing LLC.  
All Rights Reserved.

# “Tangled Views”









# MarcFrame and KBV Continued

To take us out of the bounds of MARC.

More intelligent handling of various shapes of RDF. (Differences in vocabulary, data granularity and identity; types vs. “type-like”.)

Normalizing and reducing the MARC exports by:

1. Phasing out redundant or costly details.
2. Deriving the rest (e.g. GenreForm, Content, Media, Carrier).