Technical innovations in the 20th century (a selection)

1900: The Zeppelin
1909: Instant coffee
1913: Crossword puzzles
1918: Fortune cookies
1921: The first robot
1934: The game Monopoly
1935: The first canned beer
1941: Z3, the first computer controlled by software
1945: Vannevar Bush proposes hypertext
1948: The frisbee
1957: The Fortran computer language
1959: The microchip
1962: The audio cassette
1966: The MARC format
1967: The first handheld calculator
1968: ARPAnet (first version internet)
1972: The word processor
1973: Ethernet (local computer network)
1974: Post-it notes
1984: Apple Macintosh
1990: The World Wide Web and Internet protocol (HTTP) and HTML
1995: DVD
Thus it happened 14 years ago...

http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/#_

MARC Must Die

By Roy Tennant

When MARC was created, the Beatles were a hot new group and those of us alive at the time wore really embarrassing clothes and hairstyles. Computers were so large, complex, and expensive that it was ludicrous to think that you would one day have one in your home, let alone hold one in the palm of your hand. Although age by itself is not necessarily a sign of technological obsolescence (how much has the wooden pencil improved in the last 40 years?), when it comes to computer standards it is generally not a good thing.

The very nature of the MARC (machine-readable cataloging) record is, to some degree, an anachronism. It was developed in an age when memory, storage, and processing power were all rare and expensive commodities. Now they are ubiquitous and cheap.
It’s not the lack of initiatives

**Diffusion of innovations** is a theory that seeks to explain how, why, and at what rate new ideas and technology spread.

Collingridge’s dilemma

"When change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult, and time-consuming."

- David Collingridge, The Social Control of Technology, 1980

https://en.wikipedia.org/wiki/Collingridge_dilemma
http://criticaluncertainties.com/2013/10/28/collingridges-dilemma/
Where can we go from here?
How do we handle change?
What do we need to leave behind?
The Road To Salvation

Visions of Improvements

To finally handle billions of records … with all variants of data.

And have intuitive user interfaces for all kinds of work.

Build the road. Walk the distance.
Libris

2008: We published our catalog as open linked data

2012: Inhouse development started to build a new system based on linked data (Libris XL)

2016: ... And we're still going
Storage

Search index

Graphical User Interface

Export

Import

Export

Export

Export

Export

Import

Import

Import

Import
Hold on a minute ...

Most import channels differ from one another.

Libraries rarely want their data the same way.

Catalogers need different user interfaces.

The data consists of a multitude of media types, master records, place holders, and 20+ years of "adaptations", none of which are compatible with anything else.
Assumption:
At least MARC is stable
Local Conventions and Deviations

Diffuse notion of identity.

Details captured in oddly differentiated structures or obscured in notes.

Redundant descriptions of related things, incessantly copied.

Lots of locally defined external rules and microsyntaxes, lacking scope.
MARC is not the real problem

Using a presentation format for data interchange is a symptom of the problem.

Encoding the knowledge of librarians as MARC is a symptom of the problem.

Making librarians think in MARC is a symptom of the problem.

There is a lack of cohesion and purpose.
“We can’t solve problems with the same kind of thinking that created them”  - Albert Einstein
Assumption: We know what we need to do
What are we trying to facilitate?

As librarians? As developers?

*What developers hear* > “We need faster horses.”

*What librarians hear* > “We have a robot unicorn!”

*What patrons want* > “Is this available as an e-book?”
What is usage?

We behave in different ways in different contexts and use data differently.

- We are all producers and consumers.

When is innovation creating value…

- We don’t know what value is until we make the data available for use.

There is a difference between needs and wants, we can’t use the same hammer for all problems.
Assumption: RDF solves everything
</record/1>
marc:data (  
    [ a marc:Bib260;  
      marc:bin260-ind1 marc:bin260-ind1-2;  
      marc:parts (  
        [marc:bin260-a "Bonnier"]  
        [marc:bin260-b "Stockholm"]  
        [marc:bin260-a "Bonnier"]  
        [marc:bin260-b "Malmö"]  
      ) ]  
  ) .
Removing Barriers

RDF *enables* interlinking, typed, structured values and extensible precision.

You are *no longer limited* to out-of-band code lists, microsyntaxes and controlled strings for identification.

You still need to *do* the change.
Adapting to change is believing in it

Relative Advantage
The degree to which an innovation is seen as better than the idea, program, or product it replaces.

Compatibility
How consistent the innovation is with the values, experiences, and needs of the potential adopters.

Complexity
How difficult the innovation is to understand and/or use.

Triability
The extent to which the innovation can be tested or experimented with before a commitment to adopt is made.

Observability
The extent to which the innovation provides tangible results.
Change – technical or organisational?

"Yes the data is messy but why can’t we fix it?"

"New rules and technical possibilities do not align, dear Alice."

"Why?"

"Oh dear look at the time, gotta run!"
Assumption: At least we can rely on standards
A Priori?

1. How do I know which vocabularies to use?
2. Are they useful?
3. What are their limits?
The most merciful thing in the world is the inability of the human mind to correlate all its contents.

— H.P. Lovecraft. The Call of Cthulhu
We’re Not Done Yet

We need our domain experts to get on board.

People are reluctant to get on board a moving object.
Managing expectations

“Oh, but I thought the new system would fix this?”

“What do you mean, I have to change the way I work?”
Assumption: We know what we are describing
Haddocks' Eyes

A poem sung by The White Knight to Alice in Lewis Carroll's *Through the Looking-Glass*.

The song's name is called **Haddocks' Eyes**
The song's name is **The Aged Aged Man**
The song is called **Ways and Means**
The song is **A-sitting on a Gate**
Common definitions

“The closer conscious subjects stick to **common words, idioms, phrasings, and topics**, the more easily others can **surmise their meaning**; the further they stray from common expressions and topics, the wider the **variations in interpretations**. This suggests that sentences do not have meaning intrinsically; there is not a meaning associated with a sentence or word, **they can only symbolically represent an idea**.”

https://en.wikipedia.org/wiki/Pragmatics
Framing The Entity

Create an outline with context.

This is a translation of The Hobbit.

This is a movie adaptation, starring Martin Freeman.

It’s part of a collection called “My Favorite Fantasy Books”.

Rough and lossy data is fine, as long as it’s created with intent.
Identifying a thing

How do we identify something when we don’t know its essence?

Knowing what it is, is to know where it is. In its context.

X marks the spot.
BIBFRAME 2

Adapted to experience.

1. **No more Authority** (that’s about trusting a description)
2. **No more Annotation** (that’s a specialization for external data)
3. **No more Authorized Access Points** (that’s about presentation)
BIBFRAME (“vs” RDA)

+ Simple abstraction: Instance (“carrier”) and Work (“content”)
+ Concrete subclasses for more understandable data
+ Relatively small core – delegates more to other vocabularies (e.g. RDA value lists)
+ Closer to entities hidden in MARC21

- Closer to entities hidden in MARC21
- No mappings to other vocabularies (DC terms, Schema.org)
- Still some overly specific and/or string-based terms in various places (could be marked as explicit legacy)
- Work/Instance dichotomy
Does BIBFRAME (2) Solve The Usage Problem?

Doubtfully in its entirety.

But it raises the international library platform to a level where such problems *can* be solved.
HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION: THERE ARE 14 COMPETING STANDARDS.

14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES. YEAH!

SOON:

SITUATION: THERE ARE 15 COMPETING STANDARDS.
Possibilities to express and understand new aspects
How Do We Decide?

We implement!

We think in RDF, and talk Turtle to each other.

We convert from and revert to MARC.

We build interfaces for browsing and editing, and logic for matching entities.

We involve systems librarians and other domain experts in the core team.

We fail and learn.
To catalogue or not to catalogue?

Improve automated workflows, for example by using real identifiers to match entities.

Improve data provenance by linking instead of copying.

Focus on the intellectual work of describing, identifying and linking relationships, not being experts in understanding enigmatic digitally stored catalog cards.

More data isn’t always better.
ESCAPE PLAN

Make sure the organization is onboard with working with RDF and vocabularies, whether it is metadata specialists or developers.

BIBFRAME 2 appears useful enough. Use it and measure results.

Transitioning is about learning and accepting to let go of the old.

Technology is not the problem.
Conclusion

Diligence in transition

Cohesion through context

Small steps
Thank you! Questions?

Don't be strangers!

markus.skold@kb.se
niklas.lindstrom@kb.se
fredrik.klingwall@kb.se