# Do MARC or `doJSON()` .

## Data model(s) of Invenio beyond 1.x

Alexander Wagner, *for* ***JOIN***$^2$

**CERN**

**Geneva, 13.10.2015**

HELMHOLTZ GEMEINSCHAFT

DESY

# Overview

- Invenio 1
- What is a book?
- Marc21
- Invenio 2
- `doJSON()`

CC BY NC SA



DESY

# Invenio up to 1.x

Invenio is a free software suite enabling you to run your own digital library or document repository on the web. The technology offered by the software covers all aspects of digital library management from document ingestion through classification, indexing, and curation to dissemination. Invenio complies with standards such as the Open Archives Initiative metadata harvesting protocol (OAI-PMH) and uses MARC 21 as its underlying bibliographic format.

CC BY NC SA



DESY

# Invenio up to 1.x

Invenio is a free software suite enabling you to run your own digital library or document repository on the web. The technology offered by the software covers all aspects of digital library management from document ingestion through classification, indexing, and curation to dissemination. Invenio complies with standards such as the Open Archives Initiative metadata harvesting protocol (OAI-PMH) and uses MARC 21 as its underlying bibliographic format.

**Marc21 basis**

was the selling point for **JOIN$^2$**

CC BY NC SA



DESY

# Marc21 = *MA*chine-*R*eadable *C*ataloging

- International Standard by the Library of Congress
- Established in 1966
- Tested and improved for nearly 50 years
- Lingua franca for bibliographic information
- Exchange format (lots of data available)
- Growing adoption due to RDA (e. g. Germany dropping MAB)

CC BY NC SA



DESY

# Marc21 = *MA*chine-*R*eadable *C*ataloging

- International Standard by the Library of Congress
- Established in 1966
- Tested and improved for nearly 50 years
- Lingua franca for bibliographic information
- Exchange format (lots of data available)
- Growing adoption due to RDA (e. g. Germany dropping MAB)

## Marc21 as Internal Format?

CC BY NC SA



DESY

# Marc21 = *MA*chine-*R*eadable *C*ataloging

- International Standard by the Library of Congress
- Established in 1966
- Tested and improved for nearly 50 years
- Lingua franca for bibliographic information
- Exchange format (lots of data available)
- Growing adoption due to RDA (e. g. Germany dropping MAB)

## Marc21 as Internal Format?

Not necessary, however, seemless integration required

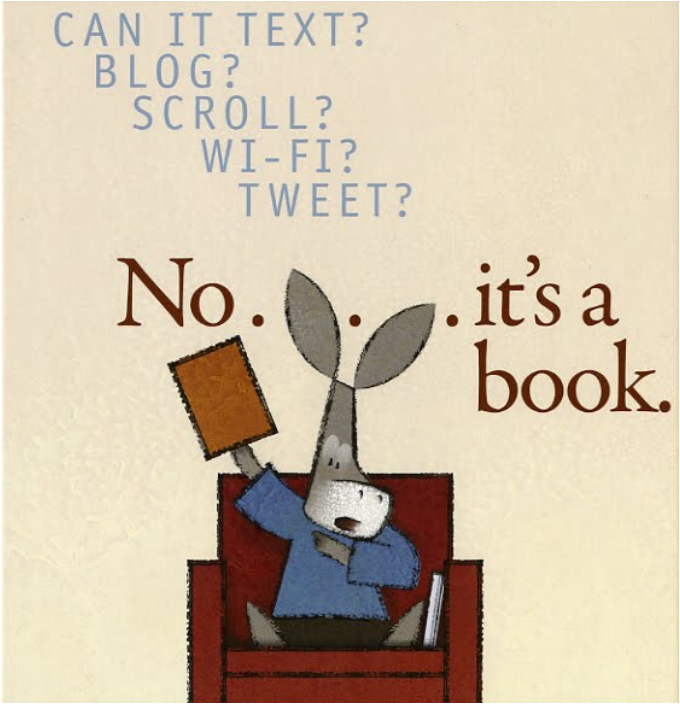
- easy, complete ingestion
- easy, complete export

CC BY NC SA



DESY

# What is a book?



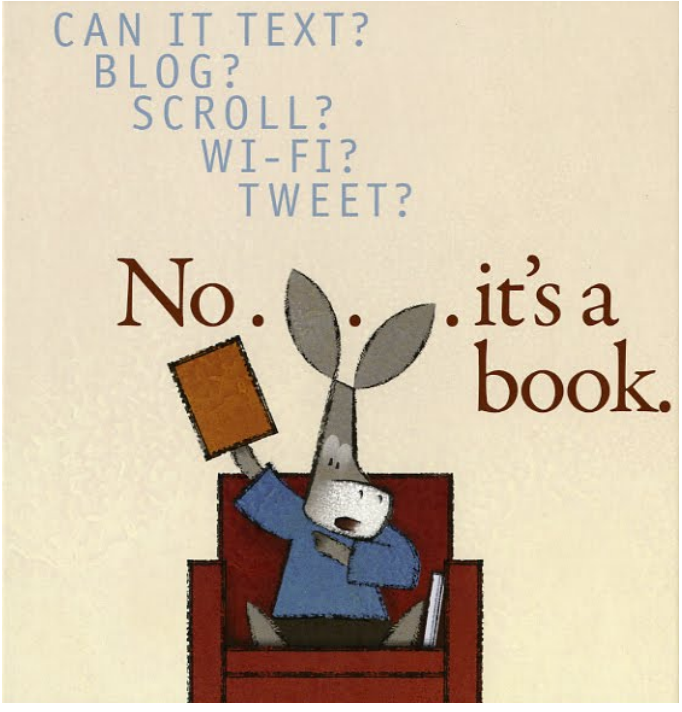

from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

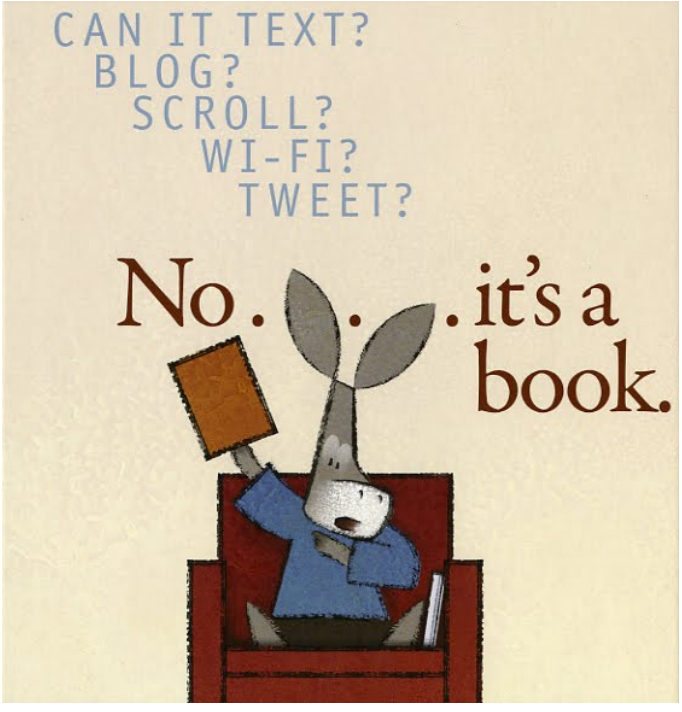
**Books are simple objects**

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

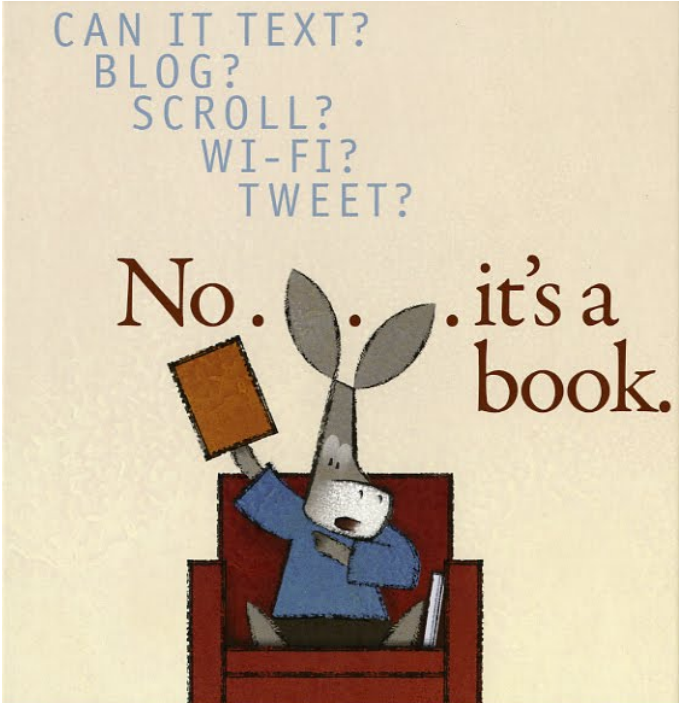
**Books are simple objects?**

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

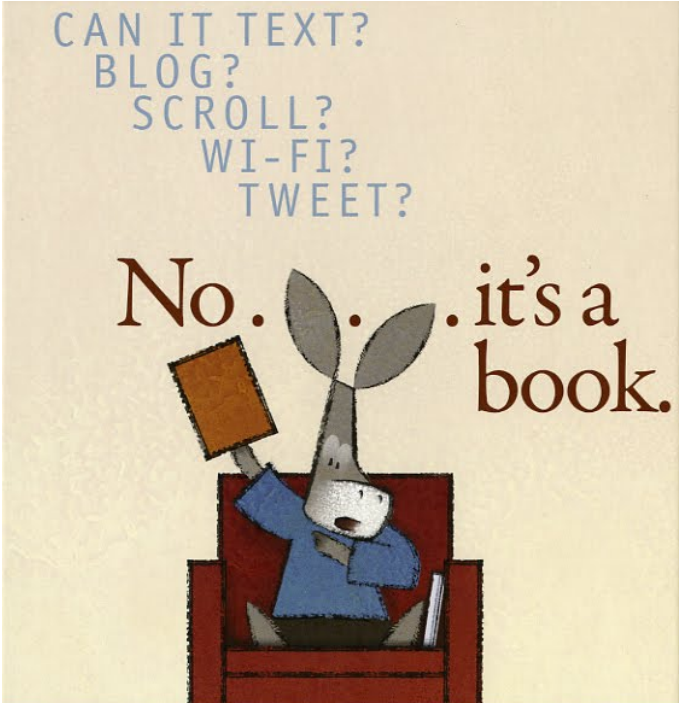
> All books have a title

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

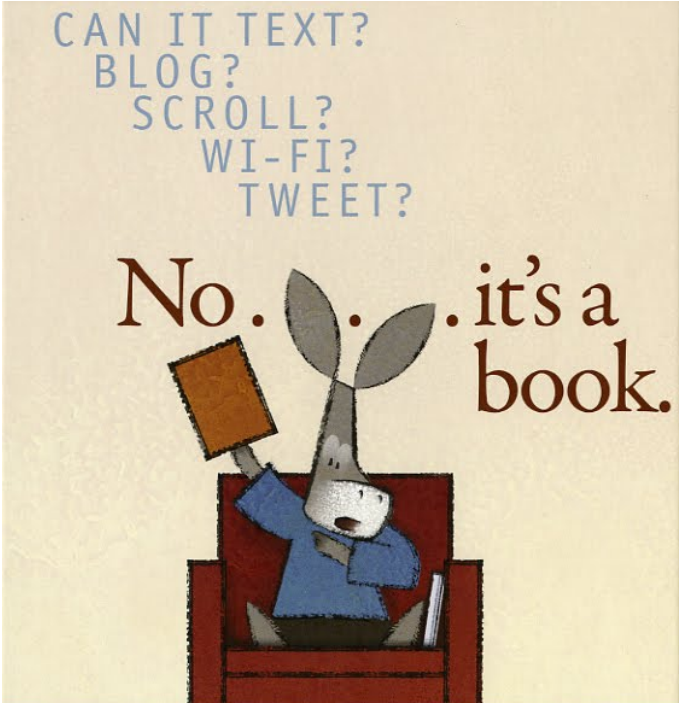
> All books have a title?

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

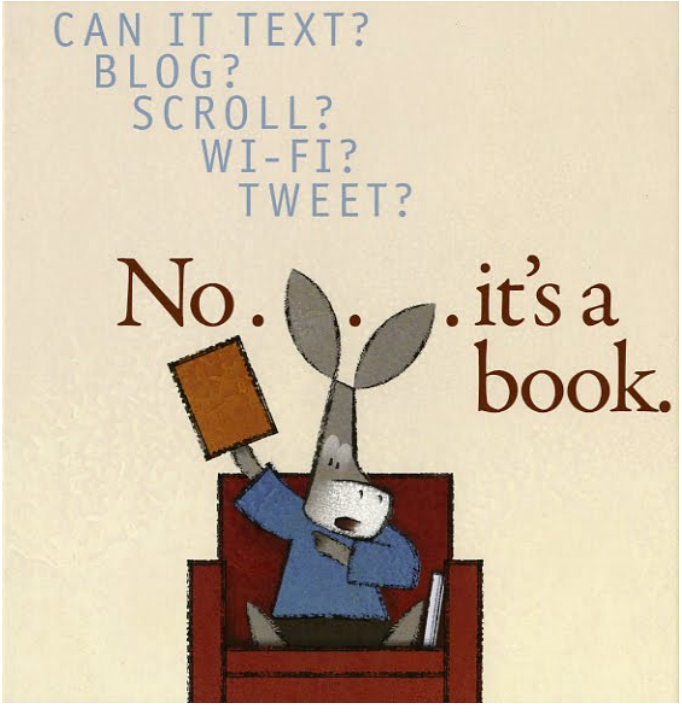
- All books have a title?
- All books have an author

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

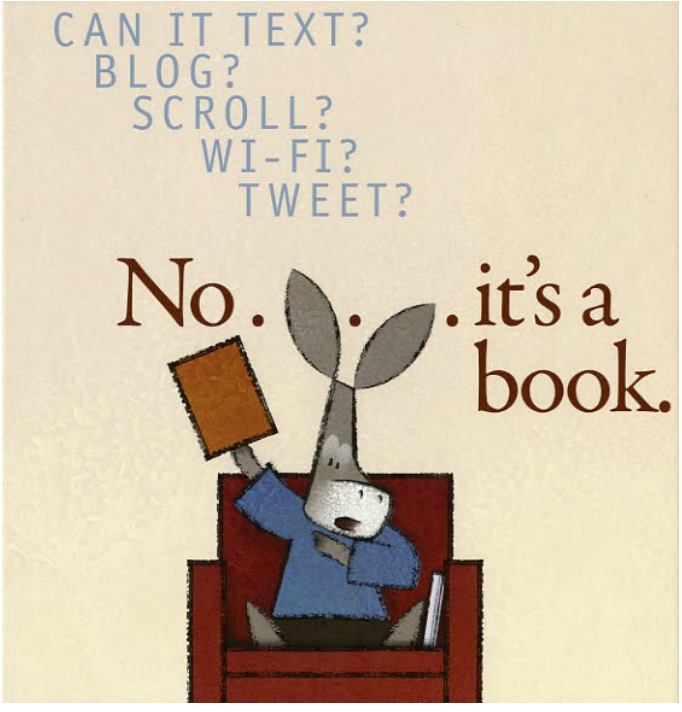
- All books have a title?
- All books have an author or editor

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

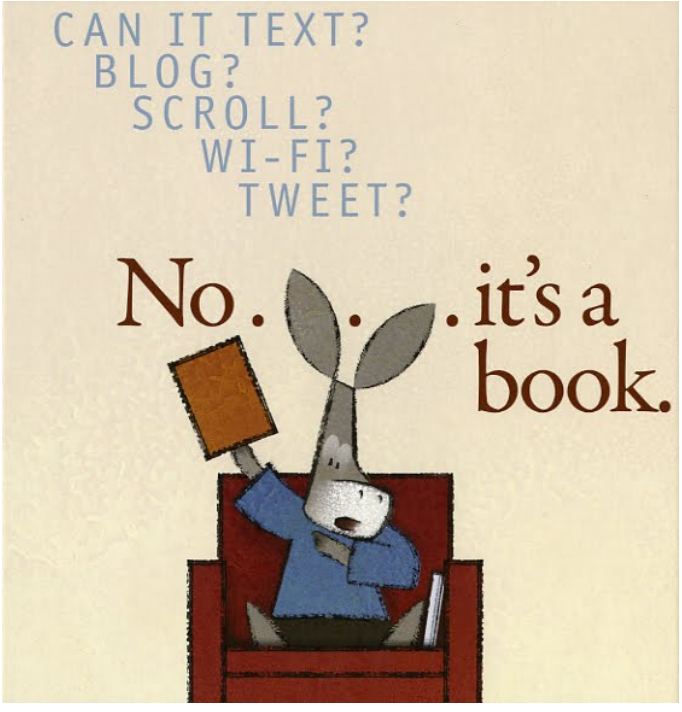
- All books have a title?
- All books have an author or editor?

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

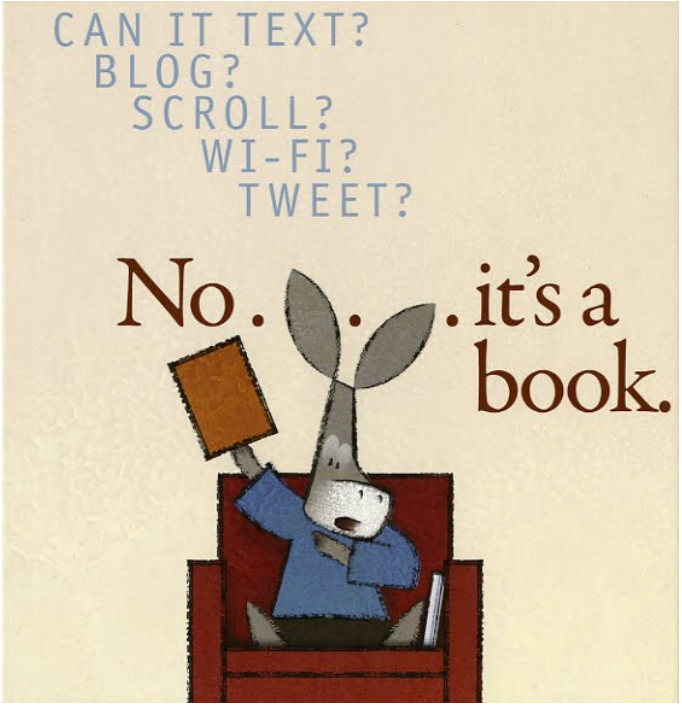
- All books have a title?
- All books have an author or editor?
- All books have an ISBN

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

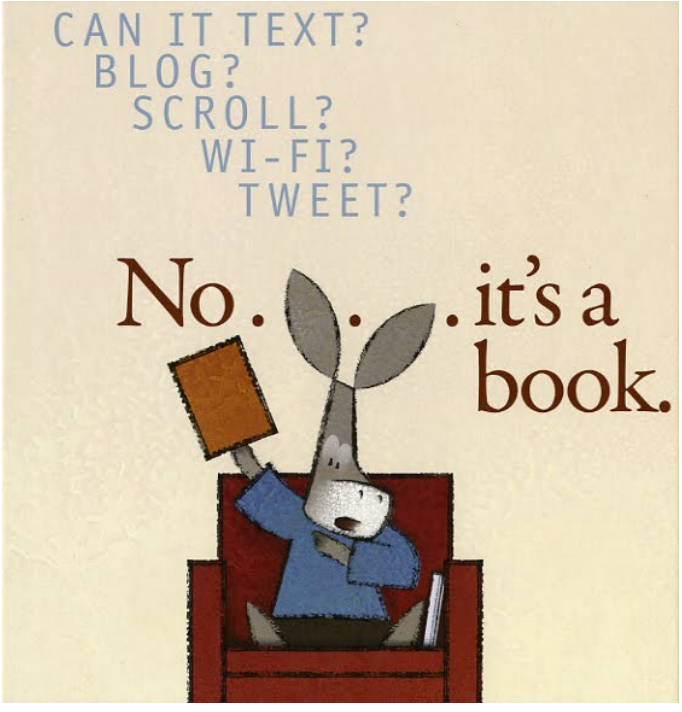
- All books have a title?
- All books have an author or editor?
- All books have an ISBN?

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

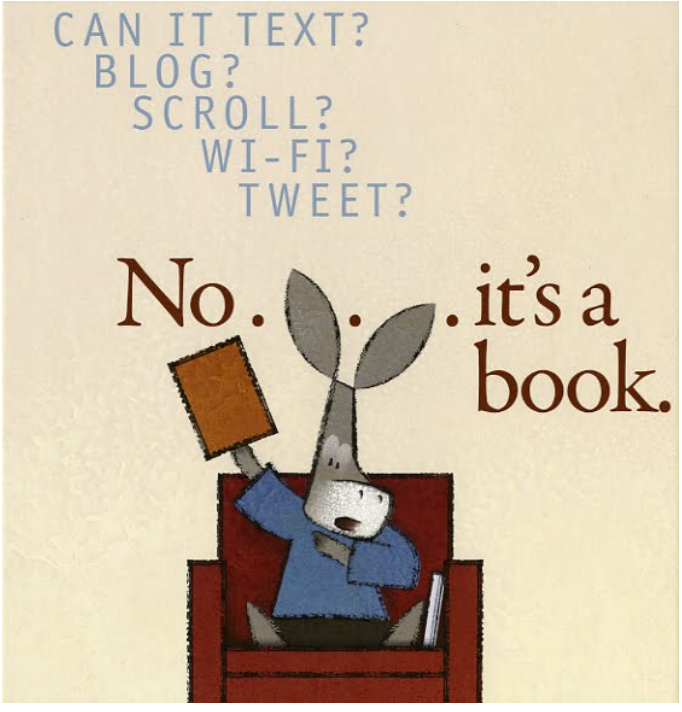
- All books have a title?
- All books have an author or editor?
- All books have an ISBN?
- All books have a publisher

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

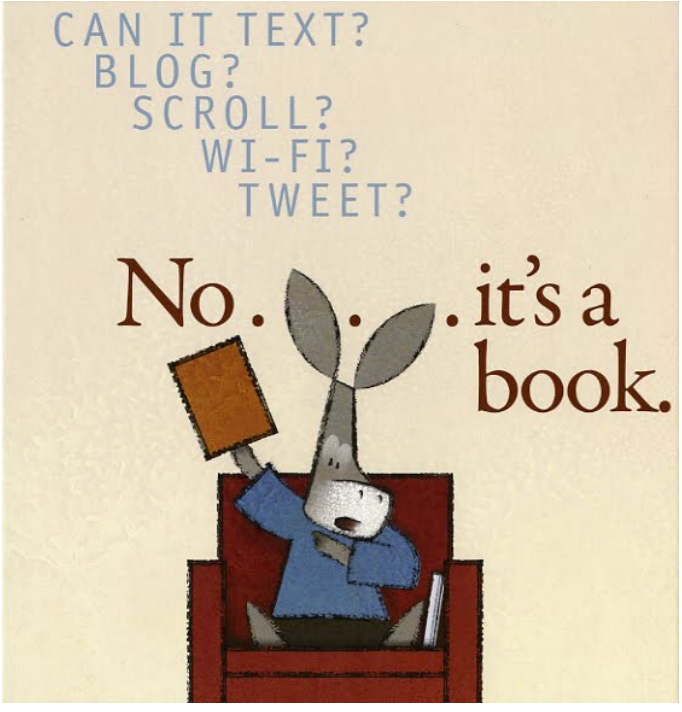
- All books have a title?
- All books have an author or editor?
- All books have an ISBN?
- All books have a publisher?

CC BY NC SA



DESY

# What is a book?




from *Lane Smith*: "It's a Book", MacMillan 2010, also on YouTube

**Books are simple objects?**

- All books have a title?
- All books have an author or editor?
- All books have an ISBN?
- All books have a publisher?

It has to be simple. You know books. Don't you?

CC BY NC SA



DESY

# This is a book…

Sächsische Akademie der Wissenschaften zu Leipzig

Herzog August Bibliothek Wolfenbüttel

## Die Deutsche Akademie des 17. Jahrhunderts Fruchtbringende Gesellschaft

Kritische Ausgabe der Briefe, Beilagen und Akademiearbeiten (Reihe I), Dokumente und Darstellungen (Reihe II)

Im Auftrag der Herzog August Bibliothek herausgegeben von Martin Birchner und Klaus Conermann

Reihe I, Abt. A: Köthen, Abt. B: Weimar, Abt. C: Halle

Reihe II, Abt. A: Köthen, Abt. B: Weimar, Abt. C: Halle

# This is a book…

# This is a book...

Sächsische Akademie der Wissenschaften zu Leipzig

Herzog August Bibliothek Wolfenbüttel

| | |
|---|---|
| **PPN:** | 741186039 Zitier |
| **Titel:** | Die Deutsche Akademie des 17. Jahrhunderts Fruchtbringende **Gesellschaft** : kritische Ausgabe **der Briefe,** Beilagen und Akademiearbeiten (Reihe I), Dokumente und Darstellungen (Reihe II) / begr. von Martin Bircher und Klaus Conermann. Im Auftr. **der** Sächsischen Akademie **der** Wissenschaften zu Leipzig, in Kooperation mit **der** Herzog-August-Bibliothek Wolfenbüttel hrsg. von Klaus Conermann<br>Reihe 1: Kritische Ausgabe **der Briefe,** Beilagen und Akademiearbeiten<br>Abt. A: **Briefe der Fruchtbringenden Gesellschaft** und Beilagen: die Zeit Fürst Ludwigs von Anhalt-Köthen 1617 - 1650 |
| **Teil:** | Bd. 6: 1641 - 1643 / unter Mitarb. von Gabriele Ball ... hrsg. von Klaus Conermann |
| **Sonst. Personen:** | Conermann, Klaus ; Ball, Gabriele *1959-* ; Bircher, Martin *1938-2006* |
| **Körperschaft:** | Fruchtbringende **Gesellschaft** |
| **Ort/Jahr:** | Berlin [u.a.] : de Gruyter, 2013 |
| **Umfang:** | 880 S. : Ill. |
| **Andere Ausgabe:** | Other edition available: Online-Ausg. |
| **Anmerkung:** | Literaturverz. S. [41] - 103 |
| **ISBN:** | 3-11-032294-3, 978-3-11-032294-1 |
| **Klassifikation:** | Dewey Decimal Classification: 430<br>Basisklassifikation: 02.01 (Geschichte **der** Wissenschaft und Kultur) ⌘ \| 02.14 (Organisation von Wissenschaft und Kultur) ⌘ \| 17.98 (Textsammlungen) ⌘ \| 18.10 (Deutsche Literatur) ⌘ |
| **Mehr zum Titel:** | Inhaltstext |

GBV S·F·X

zugehörige Publikationen

(Die deutsche Akademie des 17. Jahrhunderts: Fruchtbringende Gesellschaft. Reihe I, Abt. A: Köthen, Bd. 1. Wolfenbüttel/ Tübingen 1992)

Herzog August Bibliothek - Wolfenbüttel 2014

CC BY NC SA



DESY

# Data model: a librarians viewpoint

A bibliographic data model is a major undertaking

CC BY NC SA



DESY

# Data model: a librarians viewpoint

A bibliographic data model is a major undertaking

> Books are no simple objects

CC BY NC SA



DESY

# Data model: a librarians viewpoint

A bibliographic data model is a major undertaking

- Books are no simple objects
- Libraries hold even more material (journals, articles, media, data, games, maps...)

CC BY NC SA



DESY

# Data model: a librarians viewpoint

A bibliographic data model is a major undertaking

- Books are no simple objects
- Libraries hold even more material (journals, articles, media, data, games, maps...)
- Marc21 can cover the broad range (50 years collecting usecases)

CC BY NC SA



DESY

# Data model: a librarians viewpoint

A bibliographic data model is a major undertaking

- Books are no simple objects
- Libraries hold even more material (journals, articles, media, data, games, maps...)
- Marc21 can cover the broad range (50 years collecting usecases)
- Marc21 is complex, but not needlessly so

CC BY NC SA



DESY

# Data model: a librarians viewpoint

A bibliographic data model is a major undertaking

- Books are no simple objects
- Libraries hold even more material (journals, articles, media, data, games, maps...)
- Marc21 can cover the broad range (50 years collecting usecases)
- Marc21 is complex, but not needlessly so
- Cataloguing needs several years of education (don't throw this away)
- Cataloguing is expensive (use data if available)

CC BY NC SA



DESY

# Data model: a librarians viewpoint

A bibliographic data model is a major undertaking

- Books are no simple objects
- Libraries hold even more material (journals, articles, media, data, games, maps...)
- Marc21 can cover the broad range (50 years collecting usecases)
- Marc21 is complex, but not needlessly so
- Cataloguing needs several years of education (don't throw this away)
- Cataloguing is expensive (use data if available)

"A new model" has to raise major concerns...

CC BY NC SA



DESY

# Invenio 2.x requirements

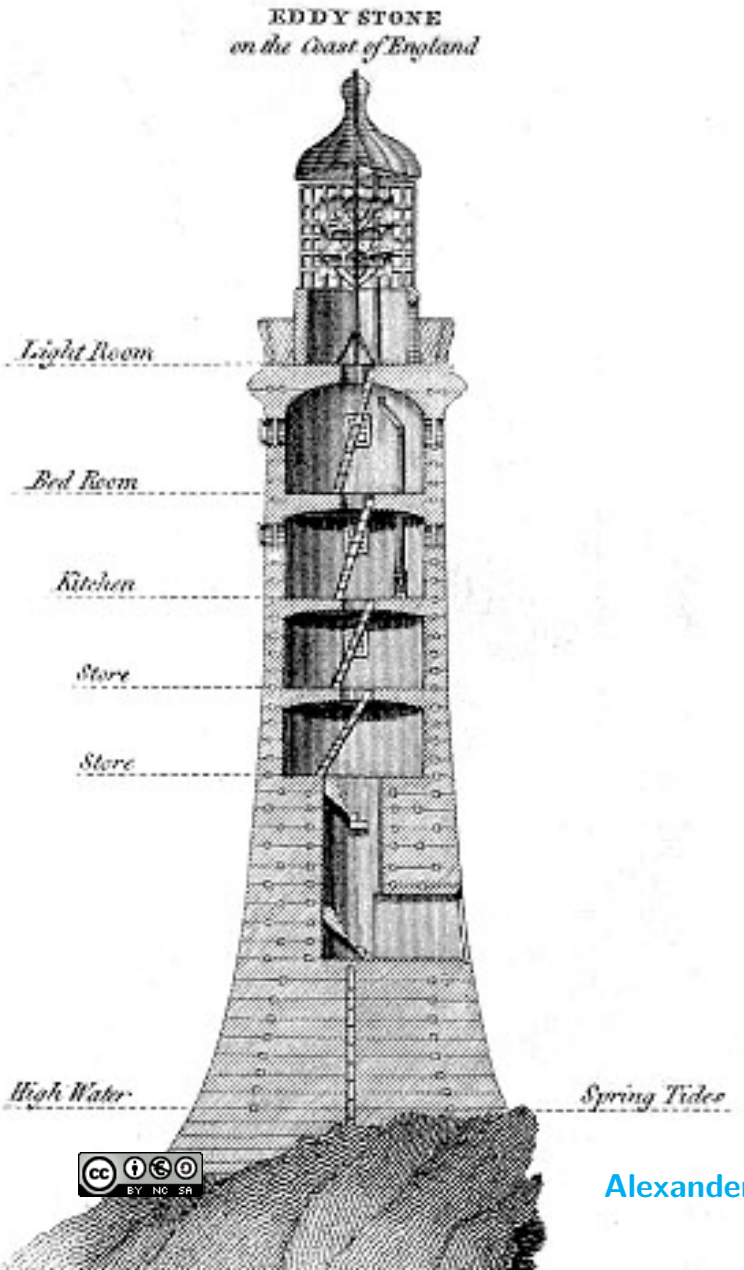
- Use OpenSource community code (e. g. programming frameworks)
- Stricter modularization
- Allow for more than one internal format (aka data model)
- Employ data validation
- Use modern data handling
- Use interlinking and expansions
- Use a modern serializtion (JSON instead of XML)
- Simple hook up with external services (Elastic Search)

CC BY NC SA



DESY

# Concerns @ JOIN$^2$

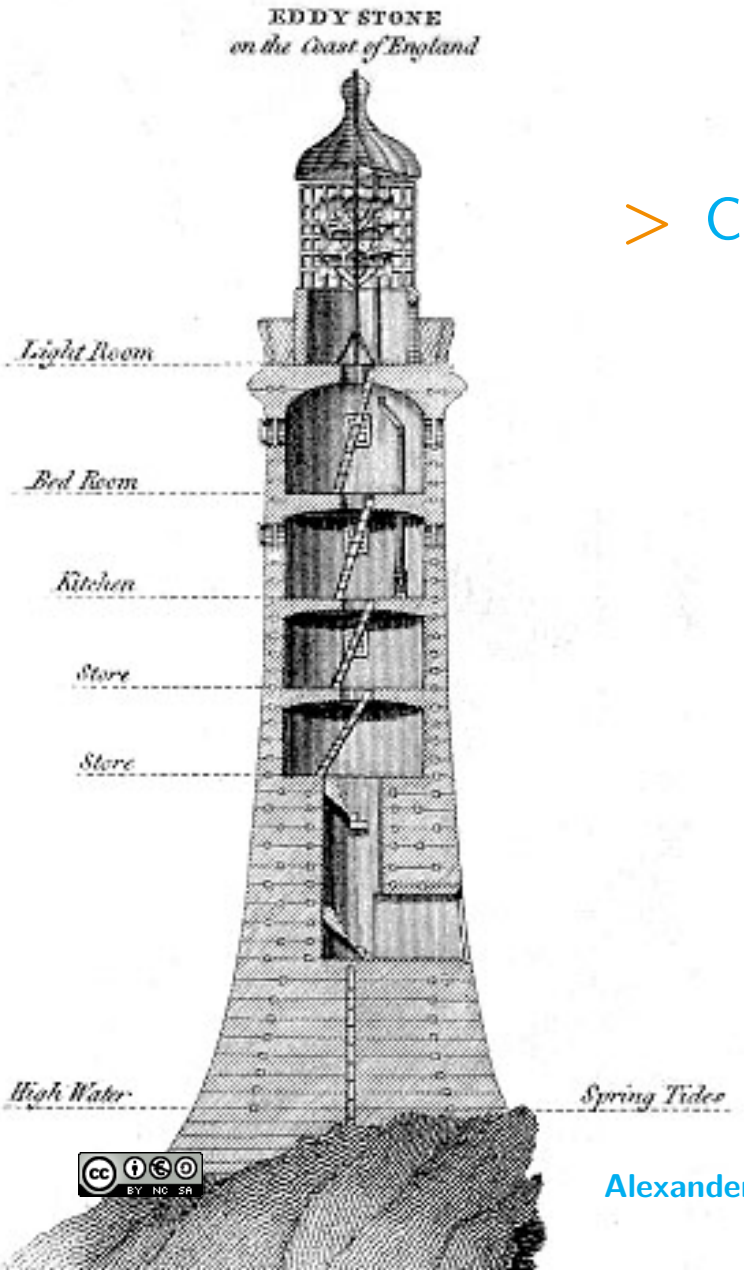



BY NC SA



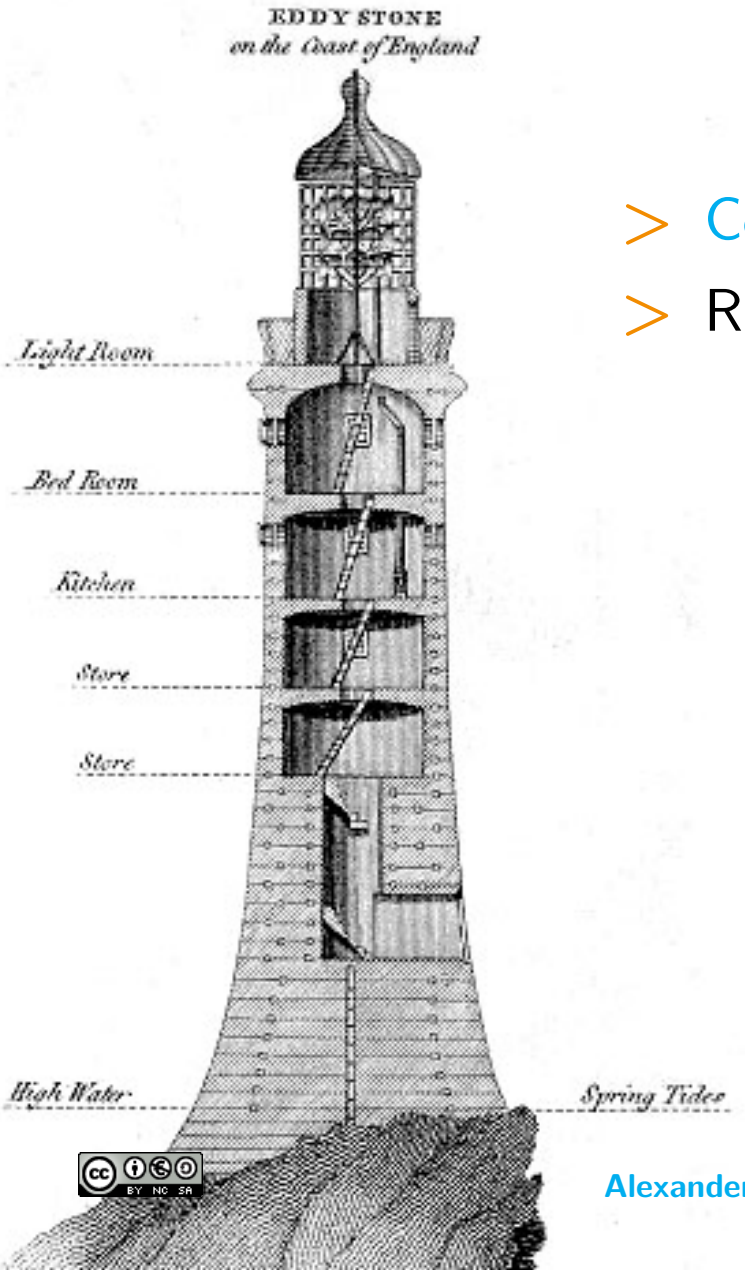
DESY

# Concerns @ JOIN$^2$



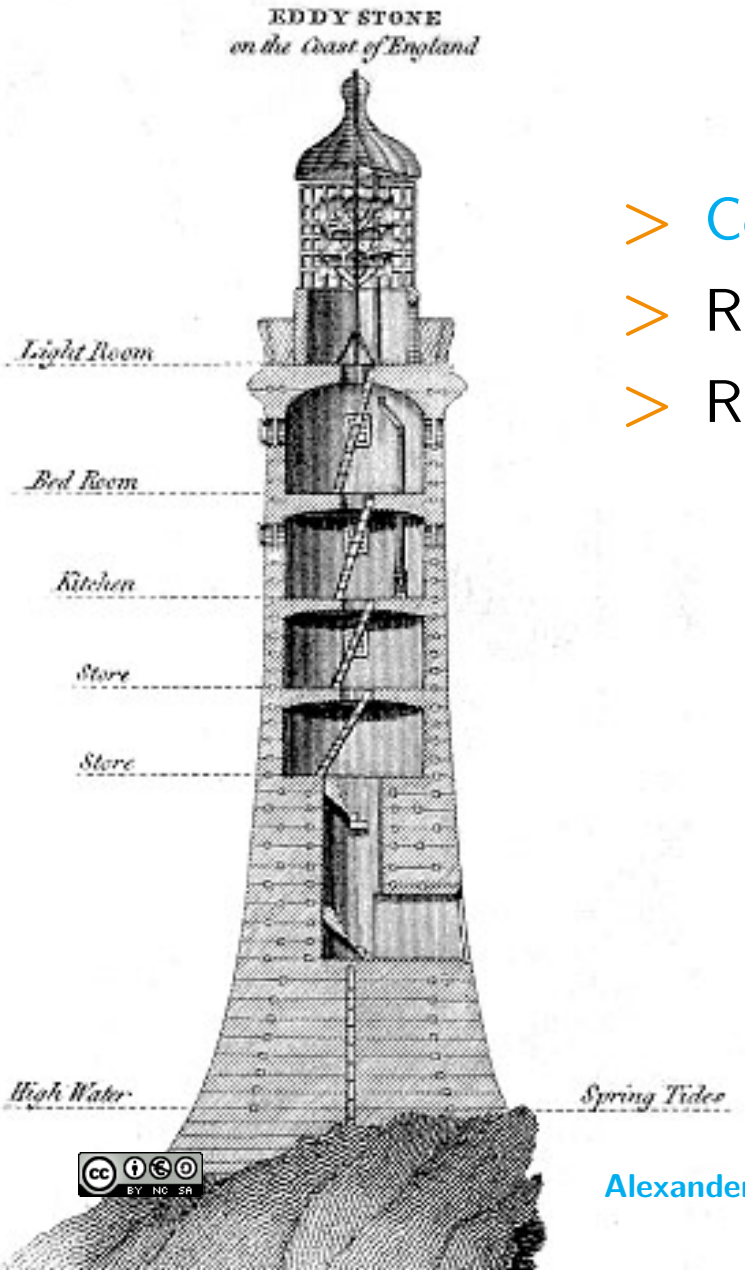

> Compatibility (1:1 mapping to and fro essential)

# Concerns @ JOIN$^2$



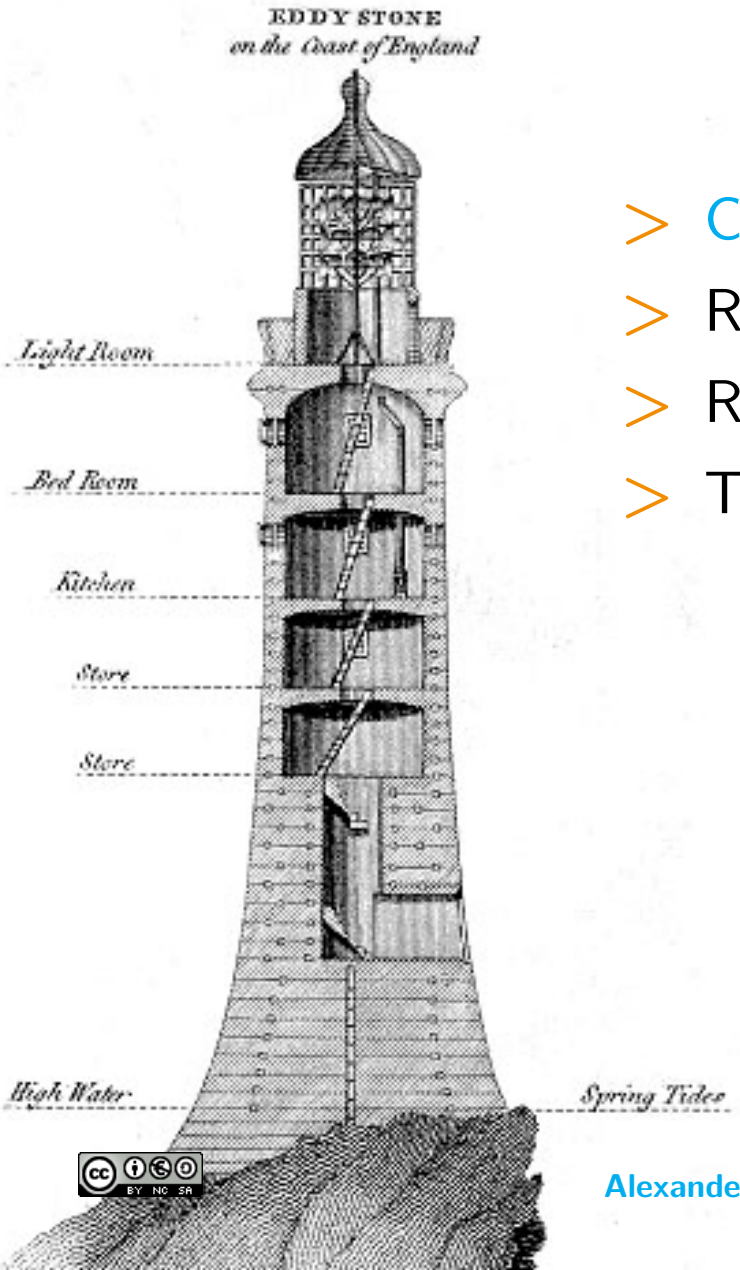

- Compatibility (1:1 mapping to and fro essential)
- Re-usage of data (e. g. ebooks, union catalogue)

# Concerns @ JOIN$^2$



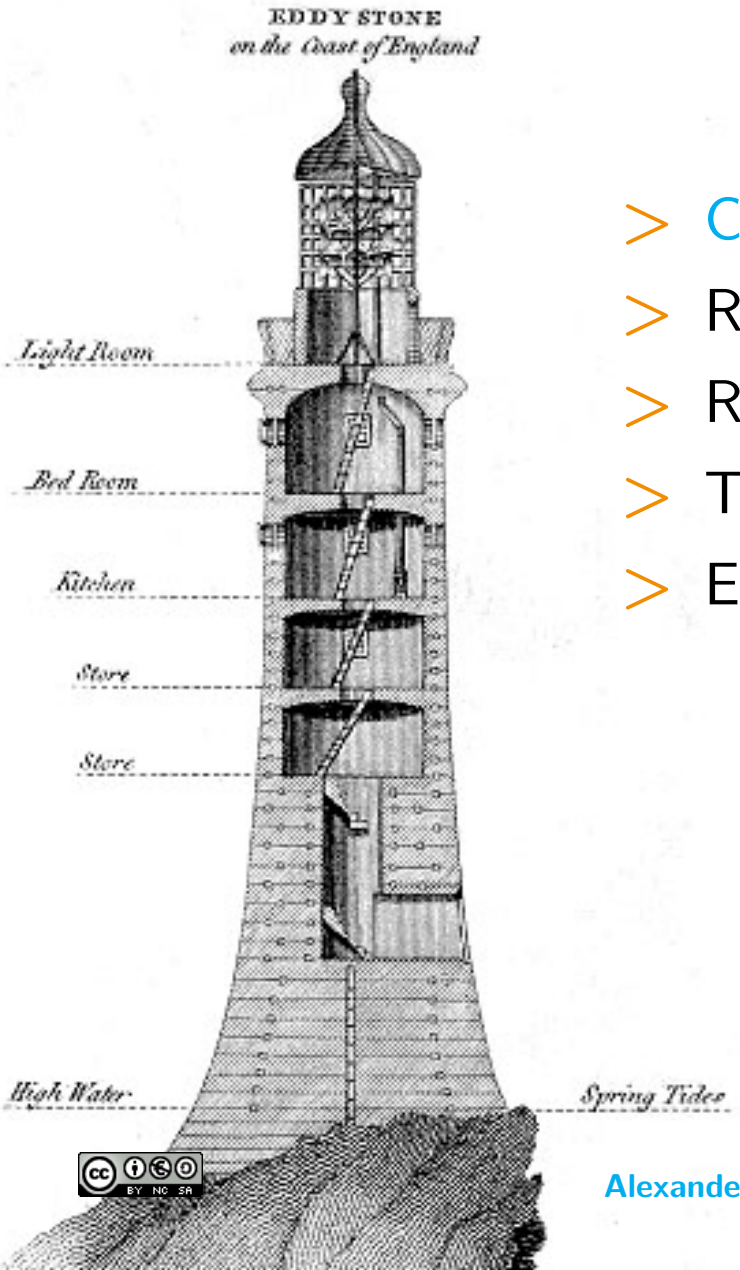

- Compatibility (1:1 mapping to and fro essential)
- Re-usage of data (e. g. ebooks, union catalogue)
- Representation of complex entities

# Concerns @ JOIN$^2$



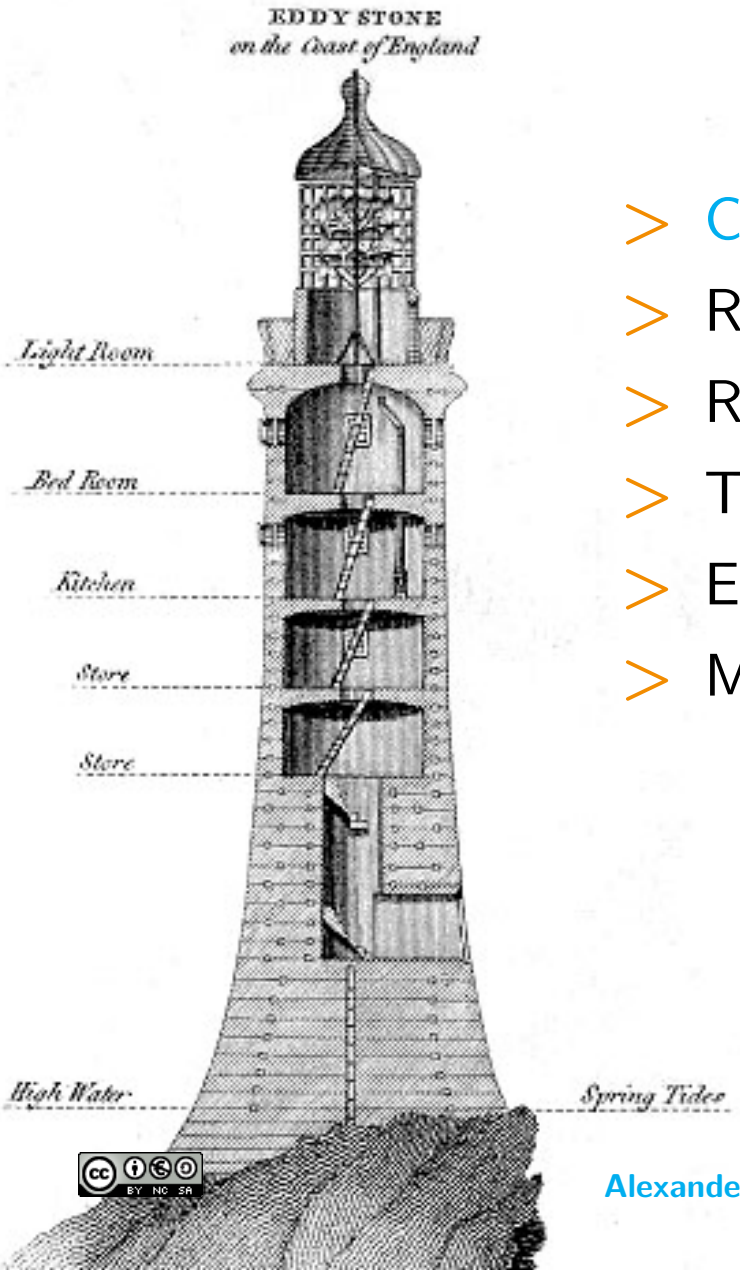

- Compatibility (1:1 mapping to and fro essential)
- Re-usage of data (e. g. ebooks, union catalogue)
- Representation of complex entities
- Training requirements for librarians

# Concerns @ JOIN$^2$



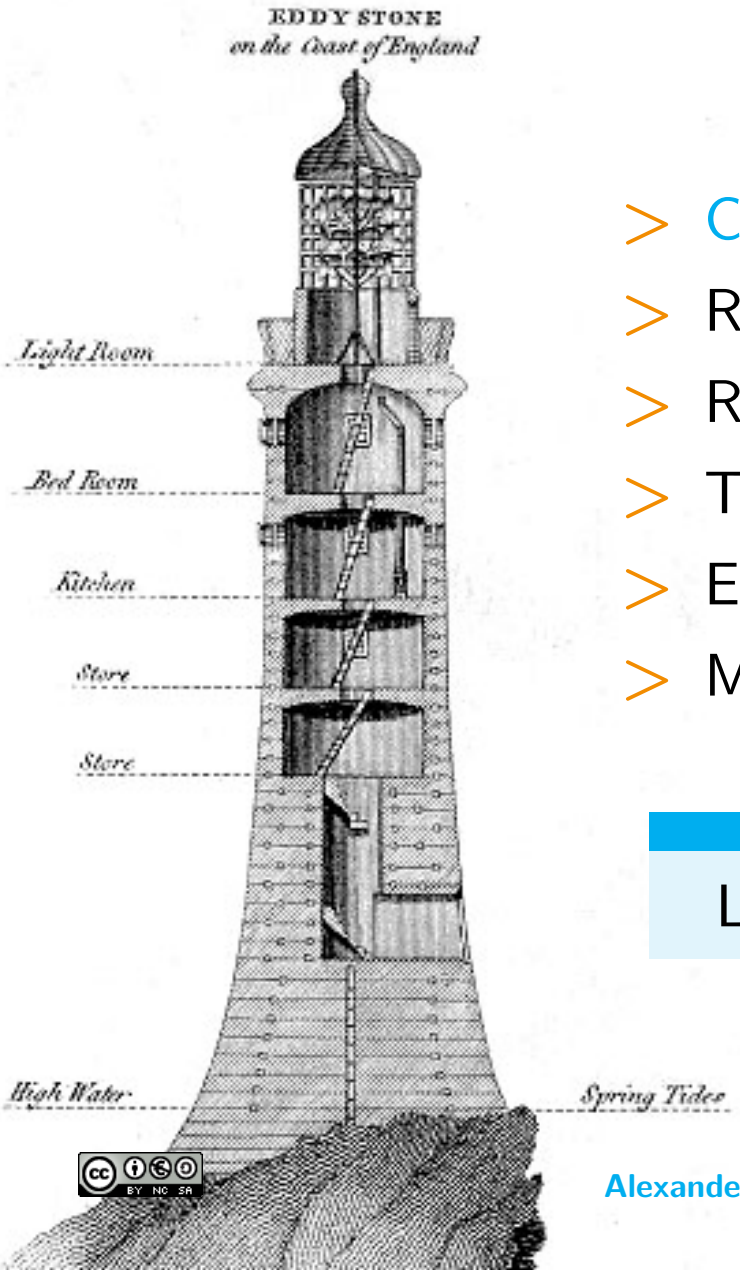

- Compatibility (1:1 mapping to and fro essential)
- Re-usage of data (e. g. ebooks, union catalogue)
- Representation of complex entities
- Training requirements for librarians
- Education (apprentices are expected to learn Marc)

# Concerns @ JOIN$^2$




- Compatibility (1:1 mapping to and fro essential)
- Re-usage of data (e. g. ebooks, union catalogue)
- Representation of complex entities
- Training requirements for librarians
- Education (apprentices are expected to learn Marc)
- Marc is compact, others are usually chatty

# Concerns @ JOIN$^2$




- Compatibility (1:1 mapping to and fro essential)
- Re-usage of data (e. g. ebooks, union catalogue)
- Representation of complex entities
- Training requirements for librarians
- Education (apprentices are expected to learn Marc)
- Marc is compact, others are usually chatty

Living on a light house is not desirable

# Addressing librarians concerns – `doJSON()`

- Marc21 available as one datamodel (serialized in JSON instead of XML)
- Covers all fields, subfields and indictators to the letter
- `doJSON()` and `undoJSON()` (1:1 mapping to Marc21)

CC BY NC SA



DESY

# Addressing librarians concerns – `doJSON()`

- Marc21 available as one datamodel (serialized in JSON instead of XML)
- Covers all fields, subfields and indictators to the letter
- `doJSON()` and `undoJSON()` (1:1 mapping to Marc21)

**Fellow librarians!**

Do check mapping for Marc21!

Marc21 strict should be perfect. A liberal scheme can then be derived.

`doJSON()` Marc Schema on github

CC BY NC SA



DESY

# Addressing librarians concerns – `doJSON()`

Still missing:

- Marc Authority (issue #7)
- Marc Holdings (issue #21)
- `bibedit` (check with iNSPIRE)

CC BY NC SA



DESY

# Addressing librarians concerns – `doJSON()`

Still missing:

- Marc Authority (issue #7)
- Marc Holdings (issue #21)
- `bibedit` (check with iNSPIRE)

Please contribute!

CC BY NC SA



DESY

# Addressing librarians concerns – `doJSON()`

Still missing:

- Marc Authority (issue #7)
- Marc Holdings (issue #21)
- `bibedit` (check with iNSPIRE)

Please contribute!

Still difficult:

- `enum` in data model (maintainability)
- "Human readable names" (longish, and every cataloguer knows 245__a, cf. #35)
- Migration of current records + history (legacy data may not validate)

CC BY NC SA



DESY

# Thank you!

Alexander Wagner
Deutsches Elektronen-Synchrotron
Central Library

Tel.: +49-40-8998-1758
alexander.wagner@desy.de

`http://library.desy.de`

Typeset by lualaTeX