



Réseau des bibliothèques de Suisse occidentale
Westschweizer Bibliotheksverbund
Rete delle biblioteche della Svizzera occidentale
Library Network of Western Switzerland

From MARC-XML to JSON-LD

A New Invenio Data Model for Bibliographic Objects and Beyond

Johnny Mariéthoz

HEG-Genève, 2016/08/04

RERO DOC

RERO DOC: the RERO Invenio Instance

- ▶ RERO digital library
- ▶ started in 2004
- ▶ 35'000 documents
- ▶ 215'000 print media issues
- ▶ 44 institutions
- ▶ content: heritage **and** scholarly documents
- ▶ based on **Invenio** 1.x with patches

RERO Customizations

- ▶ Web design
- ▶ 1st page design with content information
- ▶ purpose of **Elasticsearch**
 - ▶ hierarchical facets navigation
 - ▶ search results highlighting
 - ▶ press page
 - ▶ multilingual full text search
- ▶ **HTML templates** introduction
- ▶ document viewer **Multivio** - an e-lib.ch project
(<http://www.multivio.org>)
- ▶ visitor statistics page
- ▶ 1st JSON-LD/schema.org version

thanks to the Invenio team

RERO DOC before 2013

rero
Réseau des bibliothèques de Suisse occidentale
Wissensnetz Bibliothekenland
Rete delle biblioteche della Svizzera occidentale
Library Network of Western Switzerland

Rechercher Sourcement Informations

FR EN DE IT

Accueil

RERO DOC: nouvelle interface dès le 11.02.2013 [plus d'info]

Qu'est-ce que RERO DOC? Déposer sur RERO DOC: Pourquoi? Comment?

Qu'est-ce que RERO? Catalogue Collectif RERO

Actualités

bibliothèque numérique RERO DOC

Chercher dans 28,267 notices:

tous les champs Rechercher Liste

Conseils de recherche :: Recherche avancée

Limiter par type de document:

- Livres** (2,962)
- Journaux** (5,436)
- Thèses** (2,930)
- Mémoires** (2,189)
Mémoires niveau master (575) Mémoires niveau bachelier (1,528) Mémoires de formation continue (88)
- Revues** (21)
- Postprints** (14,121)
- Preprints** (365)
- Rapports de recherche** (314)
- Cartes géographiques** (63)
- Enregistrements sonores** (26)
- Partitions** (60)

Focaliser sur:

Naviguer par site (26,277)
Fribourg (4,885) Genève (2,384) Jura (3,483) Neuchâtel (7,058) Tessin (561)
Vaud (2,957) Valais (6,298)

Naviguer par institution/haut de page (13,155)
Universités (9,983) Hautes écoles spécialisées (1,930) Ecoles Polytechniques Fédérées (1,120) Instituts de recherche (33)

Naviguer par domaine (27,600)
Arts, Archéologie, Sciences du sport (1,205) Histoire, Géographie (2,329) Langues, Littérature (2,234) Médecine et Santé (845) Sciences sociales et naturelles (9,715)
Sciences humaines et sociales (3,690) Techniques, Sciences de l'ingénieur, Informatique (1,154)

Collections Spécifiques (8,020)
Archives, valaisannes (345) Bibliothèques et musées /Ville de Neuchâtel (8) Bulletin de la Mairie de Neuchâtel (1,372) Bulletin de la Société neuchâteloise de géographie (12) Bulletin du Cercle vaudois de botanique (458) Le Chêne Chérubique (5) Chronique librairieenne (1) Colloques et séminaires (1,000) Comptes rendus (1,000) Correspondance (1) Généalogie de familles librairieennes (148) Imprimés librairieens du 18e siècle (45) Journal officiel des adresses de la confédération (4) Journal «Corridore» (1,777) Journal «Généalogie» (1) Journal «Le Réveil» (1) Journal «Le Temps» (1) Journal du Valais (11) Journal «La Liberté» (991) Journal «L'Echo des Alpes» (81) Journal «L'Express de Neuchâtel» (194) Journal «L'Impartial» (231) Journal «L'Observateur» (177) Journal «Le Peuple de Neuchâtel» (1) «Le Gouvernement» (8) La Fédération des bibliothécaires suisses (1) La Gazette de Lausanne (1) Le Rameau des Sapins (3) Travaux de logique (1) Le Véritable messager bernois de Neuchâtel pour l'an (19) Le critique des Alpes (1) Ouvrages de référence librairieens (19) Revue historique neuchâteloise (22) Vitrines (62) Working papers MAPS (24) Working papers SEB (71)

Powered by **CDS Inventis v0.99.1** · Maintenu par doc.support@rero.ch
Dernière mise à jour : 01 mai 2013, 08:44

RERO DOC Home Page

Soumettre Personnaliser Aide | DE EN FR IT | visiteur :: identification ⓘ

rero doc
Bibliothèque numérique

Voir tout...

> Recherche avancée

35'742 objets

Cercle vaudois de botanique

Bulletin du Cercle vaudois de botanique

Fondé le 26 janvier 1949 le Cercle vaudois de botanique (CVB) est une plate-forme entre les botanistes du canton de Vaud. Articles inédits, notes concernant la floristique régionale, comptes-rendus d'excursions...

[Page de la collection »](#)

OLYMPIC

Jeux Olympiques de 1968 à Mexico

Cercle vaudois de botanique

Institutions

- > Neuchâtel (7'838)
- > Valais (6'402)
- > Fribourg (6'076)
- > Genève (5'529)
- > Vaud (4'803)
- > Jura (4'376)
- > Tessin (718)

Contenus

- > Articles (18'758)
- > Livres (6'358)
- > Mémoires (4'184)
- > Thèses (3'293)
- > Périodiques (1'712)
- > Cartes géographiques (689)
- > Rapports de recherche (606)
- > Partitions (74)
- > Enregistrements sonores (46)
- > Images, photographies, affiches (3)



Actualités



RERO DOC:
statistiques de
consultation
1.10.2015



Nouvelle collection:
Actes des Congrès de
l'Association des...

07.08.2015



OAI-PMH: mise à jour
des sets
21.04.2015

Plus

RERO DOC Search Results Page

Collections Valais

Afin de mettre en valeur ses collections patrimoniales à l'usage d'un public varié, la Médiathèque Valais-Sion alimente depuis 3 ans cette interface. Elle entreprend la numérisation de nombreux documents en mettant l'accent principalement sur 4 axes:

- La numérisation de livres anciens liés au Valais par le sujet ou par l'auteur. Il peut s'agir de livres rares provenant de la Réserve précieuse, et donc difficiles d'accès, ou de documents moins prestigieux mais très intéressants pour toute une série de recherches des milles et une facettes de l'histoire cantonale (littérature, sciences naturelles, histoire régionale, etc.)
- La numérisation de travaux d'édition plus récents parfois, avec l'accord des ayants droit si nécessaire, qui accède ainsi l'accès au domaine historique.
- La numérisation d'articles scientifiques parus dans des revues de sociétés cantonales comme les Annales valaisannes des origines à 2005
- La numérisation de numéros de revues liées au Valais

> Permalien

Trié par: Pertinence ▾ 7 résultats 1

Livres	Chancelleries et notariat dans le diocèse de Sion à l'époque de maître Martin de Sion (†1306) : étude et édition du plus ancien minutier suisse
	Annemann-Doubliez, Chantal Sion : Archives de l'Etat du Valais, 2008 (Cahiers de Valaisia ; 19) ISBN: 9782970063605 L'absence d'uniformité est ainsi apparue comme une caractéristique d'une région ALTO , soumise à différentes traditions politiques... Voir détails ALTO
	Périssel Bagnoud, Danièle Sion : Valaisia Archives de l'Etat du Valais, 2003 (Cahiers de Valaisia : Beihefte zu Valaisia ; 10) ISBN: 9782970038214 Dossier AIV, 4150-B / ALTO Étudiants ALTO 27, 28, 29, 31... Tutorat ALTO Tatara, 507 PV de la séance d u 9.6.1996...
Le Nouvelliste	
Titre de presse 110 ans, 28000 éditions	
Nouvelliste (Sion)	
Sion : Nouvelliste valaisan, 1903	
Accéder au journal «Le Nouvelliste» dans la page «Presse numérisée» ALTO	
16 mars 1992: Le Châble Tax ALTO tel. (026) 31 35 50... 4 déc. 1992: Mail à 3 ALTO 36... 3 sept. 1994: Postule Robert [] anzcek 23.00 (s) Un ami de ALTO : Félicité sur..... Voir les 3 résultats dans la page du journal «Le Nouvelliste» ALTO	

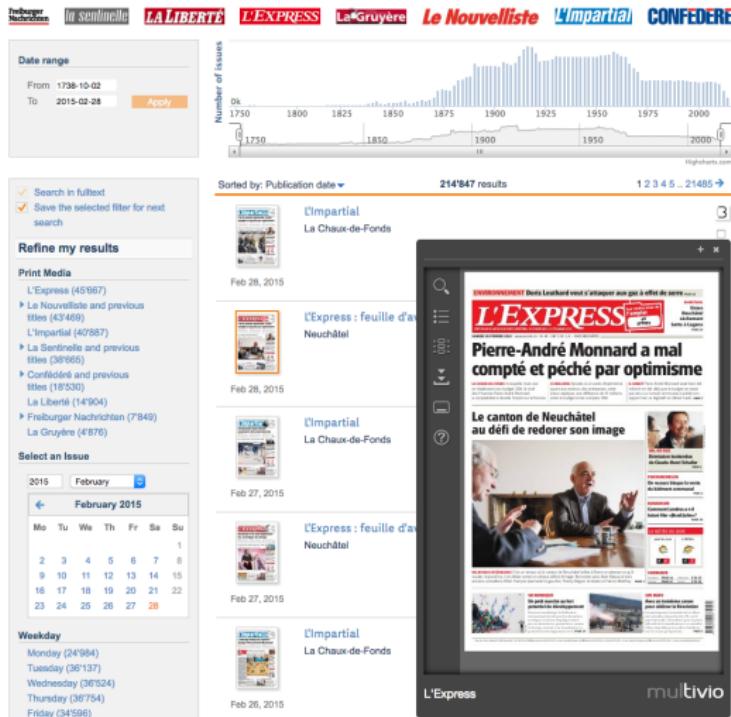
rero

Réseau des bibliothèques Suisse occidentale

HEG-Genève, 2016/08/04

RERO DOC Digitized Press Page and Multivio

Digitized Press

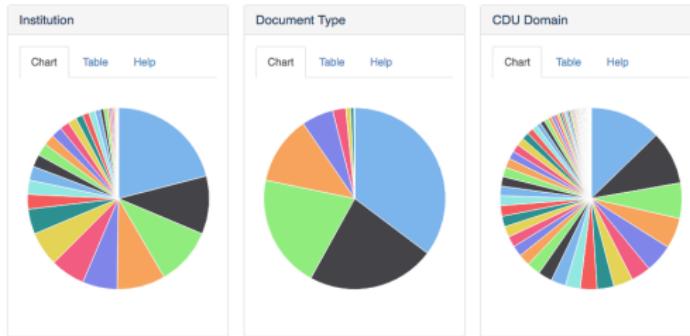


RERO DOC Visitor Statistics Page

RERO DOC - visitor statistics



Content



RERO DOC: New Challenges

- ▶ software maintenance (over Invenio versions)
- ▶ new submission interface with data import capabilities
- ▶ new services → REST API
- ▶ Invenio 3
- ▶ **Linked Open Data at RERO (<http://data.rero.ch>)**
 - ▶ at the center of our future data model
 - ▶ RERO DOC as a proof of concept
 - ▶ focus on internal and external data linking via a large use of identifiers (ORCID, etc.)
 - ▶ authority records
 - ▶ to be applied also to the **Union Catalog**

→ Invenio 3 with a new data model!

New Software

→ New Data Model?

Why Not MARC?

Is MARC Too Old?

“ By 1971, MARC formats had become the national standard for dissemination of bibliographic data in the United States (Wikipedia)

”

- 1972 C programming language is released (<http://computerhistory.org>)
- 1980 Python project is started (Wikipedia)
- 1989 Berners-Lee, Tim. "Information Management: A Proposal" (Wikipedia)
- 1990 HTML, URL, HTTP (Wikipedia)
- 1994 HTML 1.0 (Wikipedia)
- 1994 Netscape 1.0 (Wikipedia)
- 1998 Google (Wikipedia)
- 2002 Python 2.0 is released (Wikipedia)
- 2002 First CDSWare Release (Wikipedia)
- 2002 <http://json.org> started (Wikipedia)
- 2005, 2006 JSON is used by Yahoo and Google (Wikipedia)
- 2006 First Invenio Release (Wikipedia)
- 2014 RFC 7159 became the main reference for JSON's internet uses (Wikipedia)

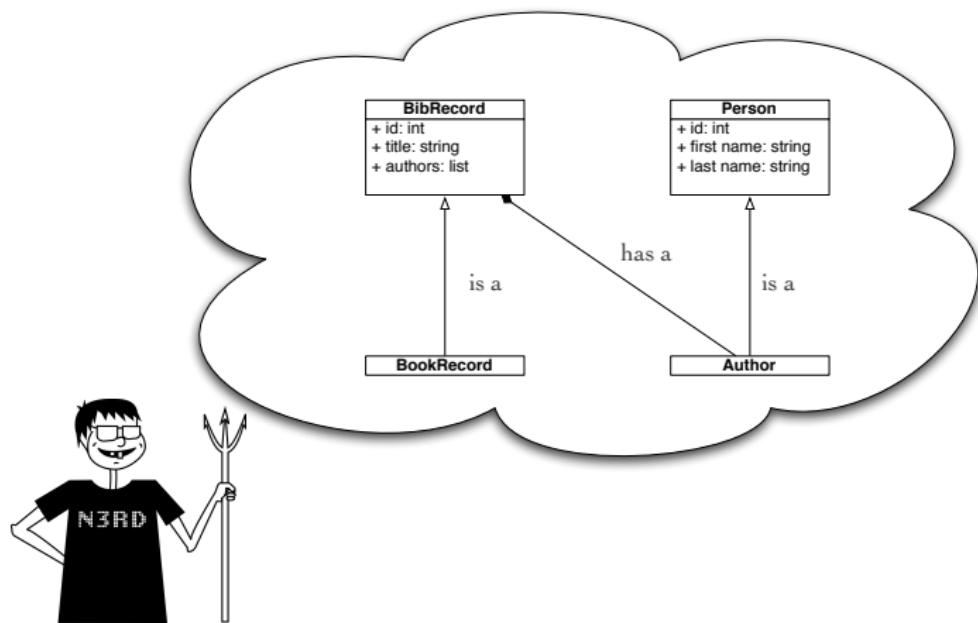
What Has Changed at the Data Level?

- ▶ THE WEB
- ▶ data handling has largely improved with new programming languages
- ▶ Web 2.0 application (client-server, services, etc.) with a lot of interactions
- ▶ emergence of Linked Open Data: everyone wants to connect to your data
- ▶ more and more exchange formats, driven by Zotero, OAI-PMH, social networks, search engines, etc.
 - ▶ → developers spend their time converting the data

MARC Was Designed for the Machines of the 70's!

What About Modern Machines?

Object Oriented Data Model



MARC Format

```
<record>
  <controlfield tag="001">1234</controlfield>
  <datafield tag="245" ind1="" ind2="">
    <subfield code="a">
      From MARC to JSON
    </subfield>
  </datafield>
  <datafield tag="100" ind1="" ind2="">
    <subfield code="a">Avram, Henriette</subfield>
  </datafield>
  <datafield tag="700" ind1="" ind2="">
    <subfield code="a">Crockford, Douglas</subfield>
  </datafield>
</record>
```



Computer Data Structures

Base Types (value)

```
title = "From MARC to JSON"  
_id = 1234  
value = 12.3
```

Computer Data Structures

List (array)

```
authors = [  
    "Henriette Avram",  
    "Douglas Crockford"  
]
```

Computer Data Structures

Dictionary (object)

```
author = {  
    "lastname": "Avram",  
    "firstname": "Henriette"  
}
```

Computer Data Structures

All Together

```
{  
  "id": 1234,  
  "title": "From MARC to JSON",  
  "authors": [  
    {  
      "lastname": "Avram",  
      "firstname": "Henriette"  
    }, {  
      "lastname": "Crockford",  
      "firstname": "Douglas"  
    }  
  ]  
}
```

Computer Data Structures

Output Format

```
{  
  "id": 1234,  
  "title": "From MARC to JSON",  
  "authors": [  
    {  
      "lastname": "Avram",  
      "firstname": "Henriette"  
    }, {  
      "lastname": "Crockford",  
      "firstname": "Douglas"  
    }  
  ]  
}
```

The JSON Format

Interesting Features

- ▶ simple: value, array, object
- ▶ easy to
 - ▶ read and write
 - ▶ share between client and server (python, javascript)
 - ▶ share (REST API)
 - ▶ work with existing libraries (Elasticsearch, Postgresql)
- ▶ can represent **any kind of object** (comments, notes, tags, libraries, collections, etc.)
- ▶ supported by many programming languages
- ▶ human readable (debug, understand)
- ▶ widely used on the **Web**
- ▶ (**too?**) flexible

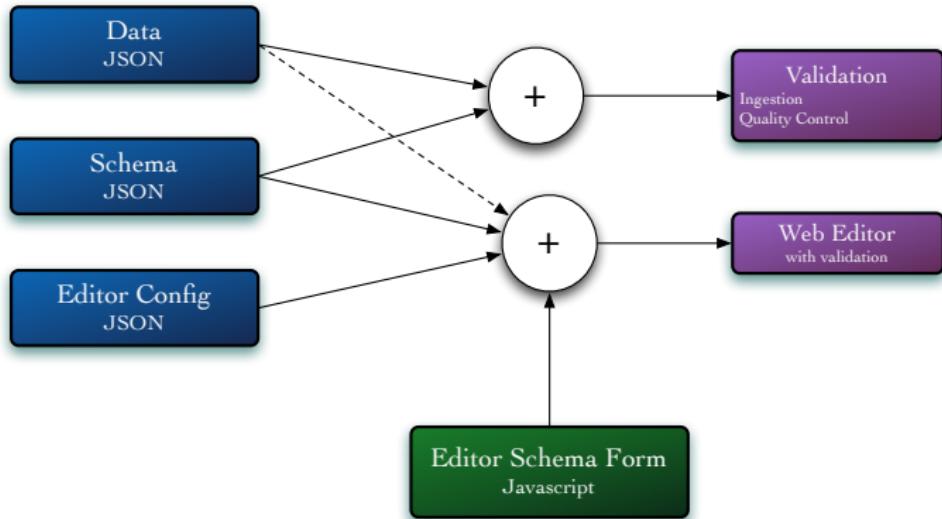
Missing Features

- ▶ standard naming (creators, authors, etc.)
- ▶ data validation
- ▶ clear format description: human and machine

→ JSON Schema

JSON Schema

The Concept



JSON Schema Advantages

- ▶ describes your existing data format
- ▶ clear, human- and machine-readable documentation
- ▶ complete structural validation, useful for
 - ▶ **automated testing**
 - ▶ validating client-submitted data

Example

Person Schema

```
{ "$schema": "http://json-schema.org/schema#",  
  "id": "/schemas/person-v1.0.0.json",  
  "title": "Person",  
  "description": "A Physical Person",  
  "type": "object",  
  "properties": {  
    "firstName": {"type": "string"},  
    "lastName": {"type": "string"},  
    "age": {  
      "description": "Age in years",  
      "type": "integer",  
      "minimum": 18  
    }  
  },  
  "required": ["firstName", "lastName"] }
```

Valid Person Data

```
{  
  "firstName": "Henriette",  
  "lastName": "Avram",  
  "age": 55  
}
```

Invalid Person Data

```
{  
  "lastName": "Avram",  
  "age": 10  
}
```

JSON Editor - Angular Form Editor

Person Schema

```
{"$schema": "http://json-schema.org/schema#",
"id":"/schemas/person-v1.0.0.json",
"title": "Person",
"description": "A Physical Person",
"type": "object",
"properties": {
    "firstName": {"type": "string"},
    "lastName": {"type": "string"},
    "age": {
        "description": "Age in years",
        "type": "integer",
        "minimum": 18
    }
},
"required": ["firstName", "lastName"]}
```

▶ Editor

Editor Configuration

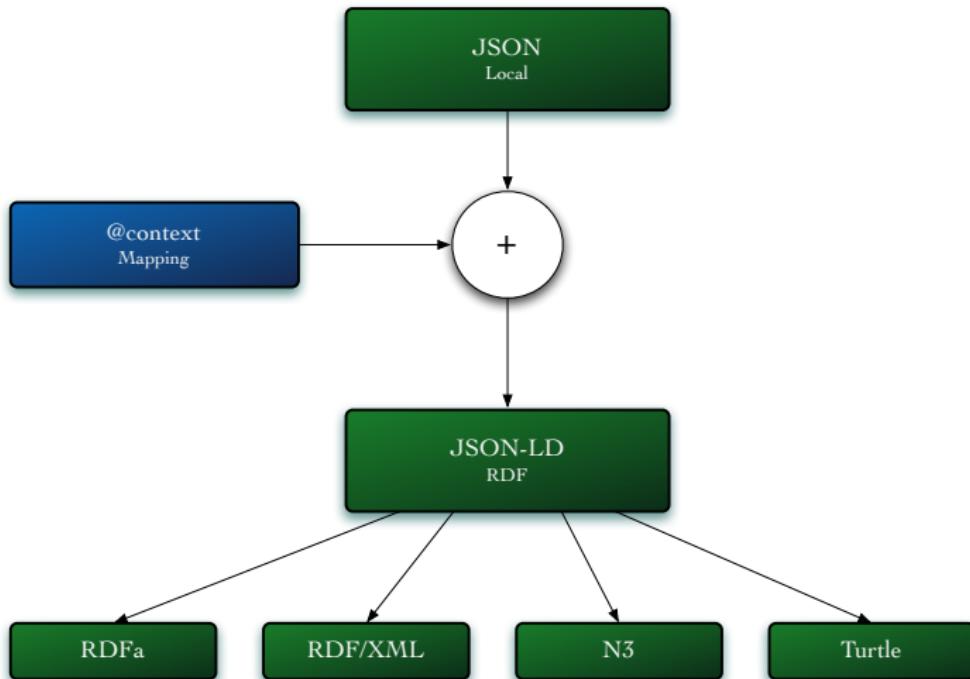
```
[{
    "key": "firstName",
    "placeholder": "please enter..."
}, {
    "key": "lastName",
    "placeholder": "please enter..."
},
"age",
{
    "key": "comment",
    "type": "textarea",
    "placeholder": "Make a comment"
}, {
    "type": "submit",
    "style": "btn-info",
    "title": "Submit"
}]
```

▶ Form ▶ Data

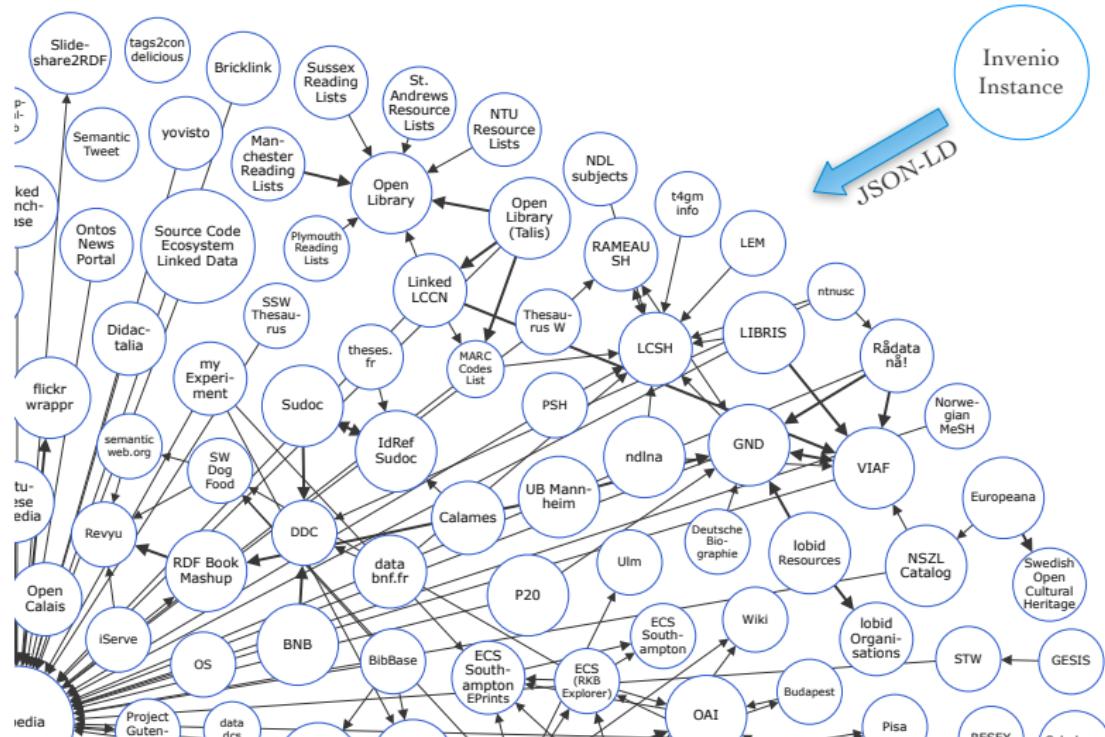
Exporting your Data

→ JSON-LD

The Concept



Your Data in the LOD Cloud



JSON Editor

Book

```
{  
  "recid": "1234",  
  
  "title": "From Marc to JSON",  
  
  "authors": [  
    {  
      "name": "Crockford, Douglas 1955-"  
    },  
    {  
      "uri": "http://viaf.org/viaf/18236820"  
    }  
  ]}
```

@context

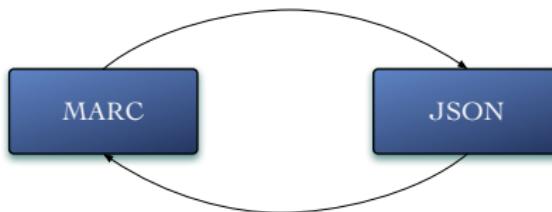
```
"@context": {  
  "dc": "http://purl.org/dc/elements/1.1/",  
  "dct": "http://purl.org/dc/terms/",  
  
  "@base": "http://doc.rero.ch/record/",  
  
  "recid": "@id",  
  "uri": "@id",  
  "name": "@value",  
  
  "title": "dct:title",  
  "authors": "dc:creator"  
}
```

▶ JSON-LD

Summary

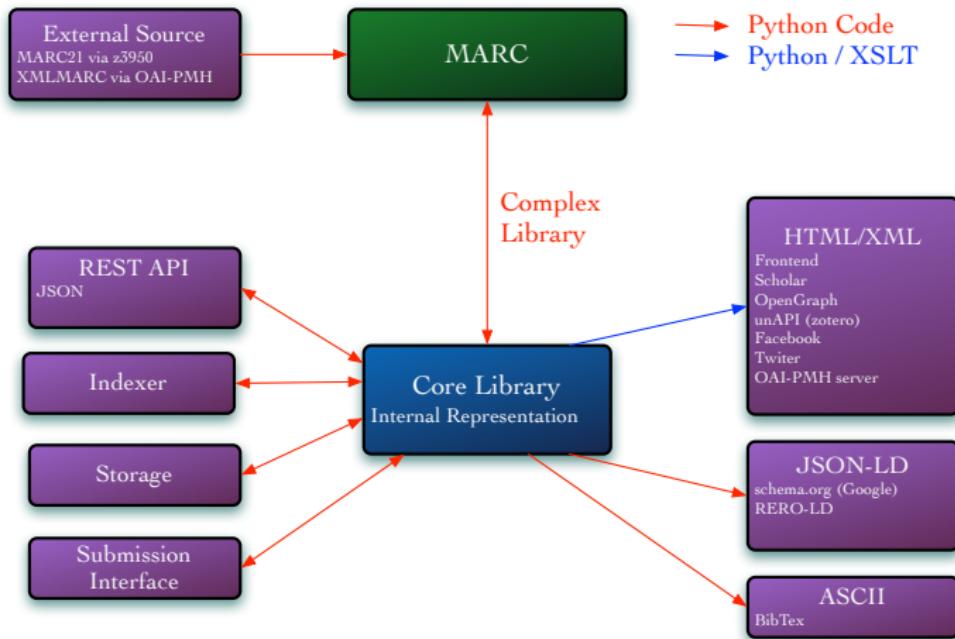
- ▶ JSON Data
 - ▶ simple
 - ▶ powerful
 - ▶ portable
- ▶ JSON-Schema Framework
 - ▶ validation
 - ▶ HTML form generation
- ▶ JSON-LD Mapping
 - ▶ lightweight data exchange

And MARC? full forward/backward compatibility

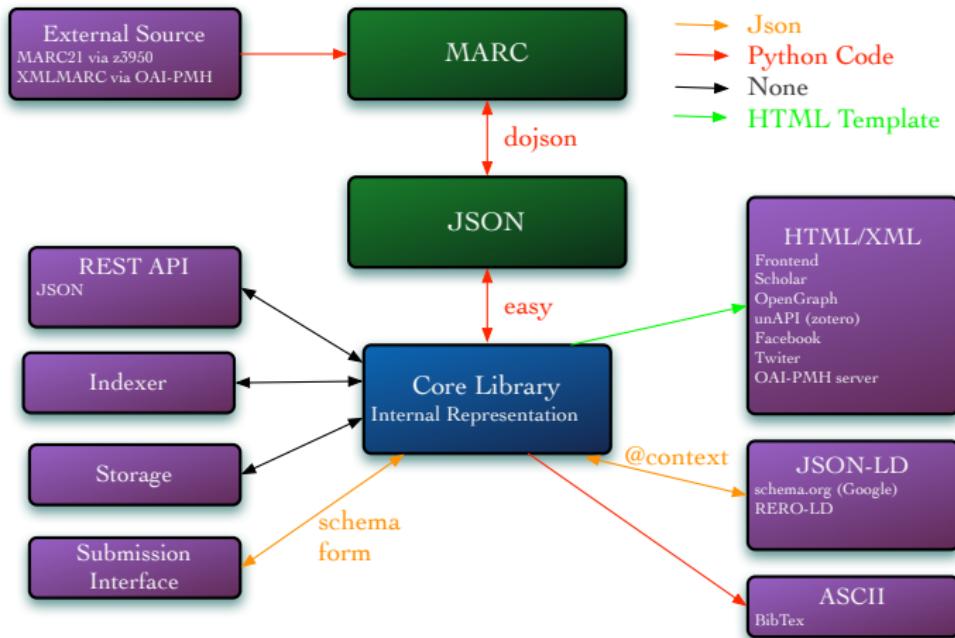


A New Data Model Based on JSON

The Current Model



The New Data Model



Conclusion

Conclusion

- ▶ Invenio 3 opens new perspectives
- ▶ JSON is obvious for the Web
- ▶ still MARC compatible
- ▶ data conversion is more affordable, robust and easier to maintain
- ▶ developers may focus on new developments
- ▶ librarians may take full control of data modeling and exchange by learning JSON-Schema and JSON-LD

References

- ▶ RERO DOC <http://doc.rero.ch>
- ▶ Invenio <http://invenio-software.org/>
- ▶ JSON-LD <http://json-ld.org/>
- ▶ JSON Schema <http://json-schema.org/>
- ▶ RERO LOD <http://data.rero.ch>
- ▶ Elasticsearch <https://www.elastic.co>
- ▶ Angular Form Editor <http://schemaform.io/>