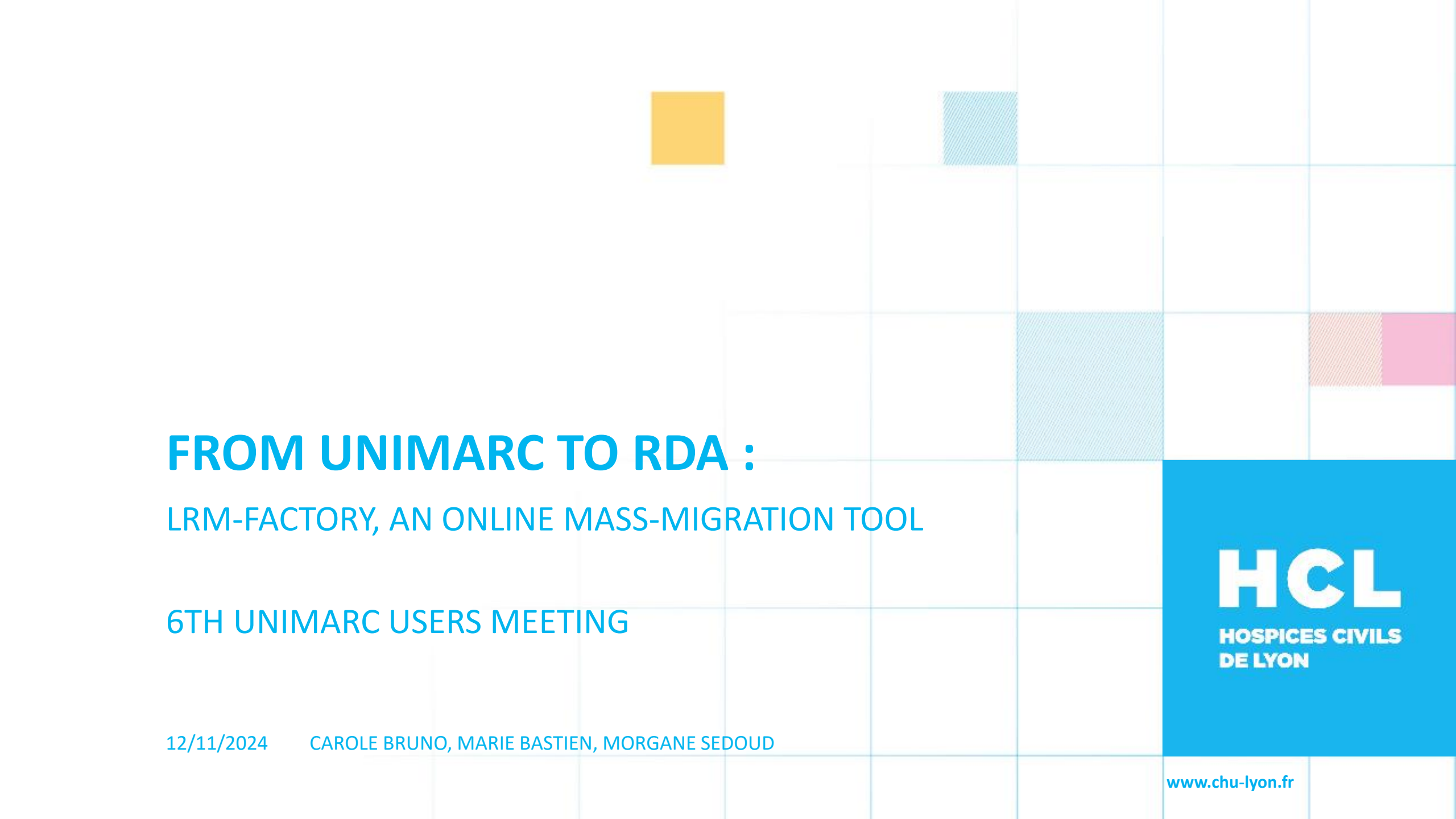




FROM UNIMARC TO RDA :

LRM-FACTORY, AN ONLINE MASS-MIGRATION TOOL

6TH UNIMARC USERS MEETING

12/11/2024 CAROLE BRUNO, MARIE BASTIEN, MORGANE SEDOUD

HCL
HOSPICES CIVILS
DE LYON

www.chu-lyon.fr

1/ Project and actors : The ACASE Consortium

- Senolys
- Tech Advantage and Syrtis
- Hospices Civils de Lyon
- The data

2/Methodology and tools

- Mapping Unimarc to RDA : From one to several mapping tables
- CoM3T Tool and Bibliographic Patterns
- Working together on sprints

3/First results and future developments

- Sprints 0, 1 and after
- Sharing our project

SEPTEMBER 2024 : IT'S A NEW SCHOOL YEAR



1/ PROJECT AND ACTORS : THE ACASE CONSORTIUM

Accès à la Culture, aux Arts et aux Sciences par les Entités

Access to Culture, Arts and Sciences through Entities

1/ PROJECT AND ACTORS : THE ACASE CONSORTIUM

4

SENOLYS

- Owns and produces CoM3T : Case-oriented MARC Metadata Migration Tool
- Joffrey Decourselle, *Migration et enrichissement sémantique d'entités culturelles*, Université de Lyon, 2018

The logo for Senolys, featuring the word "Senolys" in a bold, black, sans-serif font. The letter "l" is stylized with a vertical bar that is purple at the top and blue at the bottom, and a blue diagonal stroke.

1/ PROJECT AND ACTORS : THE ACASE CONSORTIUM

5

TECH'ADVANTAGE

- Owns and develops Syrtis (Integrated Library System) and the platform LRM-Factory
- Owns and develops the platform LRM-Factory



1/ PROJECT AND ACTORS : THE ACASE CONSORTIUM

6

HOSPICES CIVILS DE LYON

- Second university hospital of France
- Documentation experts
- Work on defining and perfecting the migration rules



1/ PROJECT AND ACTORS : THE ACASE CONSORTIUM

7

THE DATA

- 18 millions records from Abes
(Agence bibliographique de l'enseignement supérieur /
bibliographic agency of higher education)
- Funded by the French Government
operating the France 2030
investment program



2/ METHODOLOGY AND TOOLS

2/ METHODOLOGY AND TOOLS

9

MAPPING UNIMARC TO RDA : FROM ONE TO SEVERAL MAPPING TABLES

- Two examples of records : monograph/serial resource
- Mapping by Unimarc fields: explore each field (and sub-fields)
- Mapping expressed in RDF triplets.*
- Specific mappings : 7XX / 0XX / 1XX

Sujet	Prédictat	Objet
http://irm-factory.fr/ressource/RDA/Manifestation#Key	http://rdaregistry.info/Elements/m/object/P30462	http://irm-factory.fr/ressource/RDA/Nomen#Key
http://irm-factory.fr/ressource/RDA/Nomen#Key	http://rdaregistry.info/Elements/n/datatype/P80068	[Contenudu\$]^xsd:string

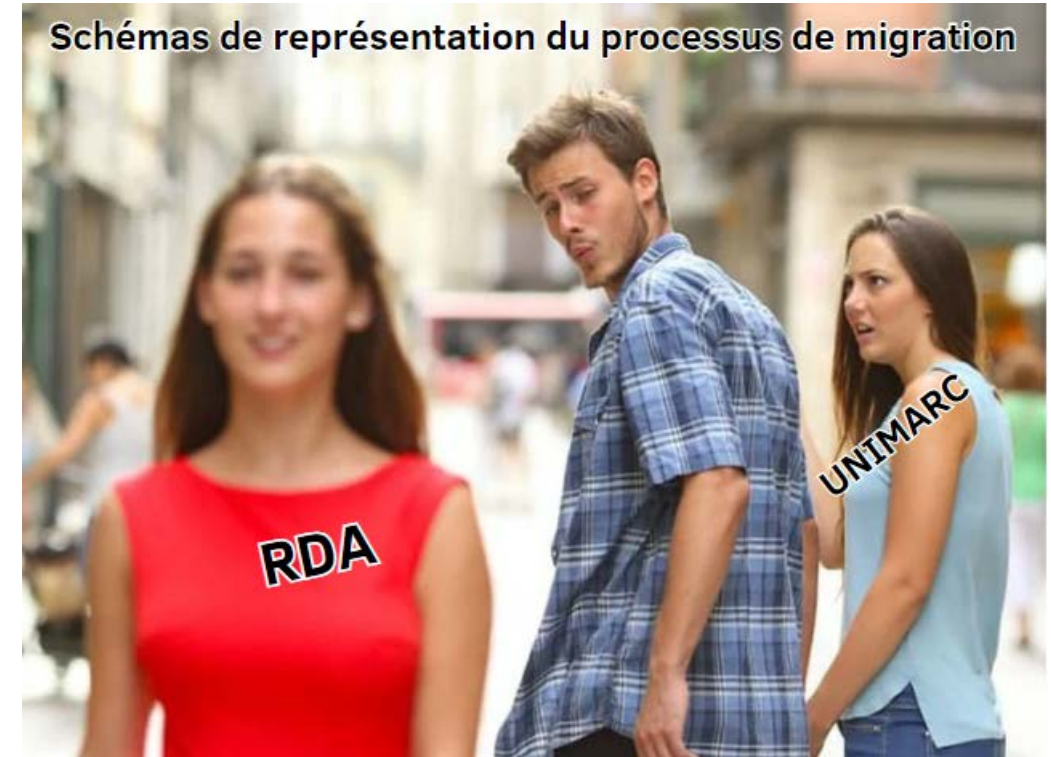
*RDF triplets mapping table

2/ METHODOLOGY AND TOOLS

10

COM3T TOOL AND BIBLIOGRAPHIC PATTERNS

- Oriented graph of bibliographic patterns / meta-model of rules : entities and properties from the target model



*Representation of the conversion process

2/ METHODOLOGY AND TOOLS

11

COM3T TOOL AND BIBLIOGRAPHIC PATTERNS

Integration process into CoM3T :

- Selection of the ten most represented fields in the incoming dataset in UNIMARC
- Mapping using our RDF conversion table
- Adjustment of the « outcoming » vocabulary and of the « incoming » vocabulary
- Addition of the rules into CoM3T : describing conditions, views and mapping
- Launch of sprints through LRM-Factory and evaluation

2/ METHODOLOGY AND TOOLS

12

WORKING TOGETHER ON SPRINTS

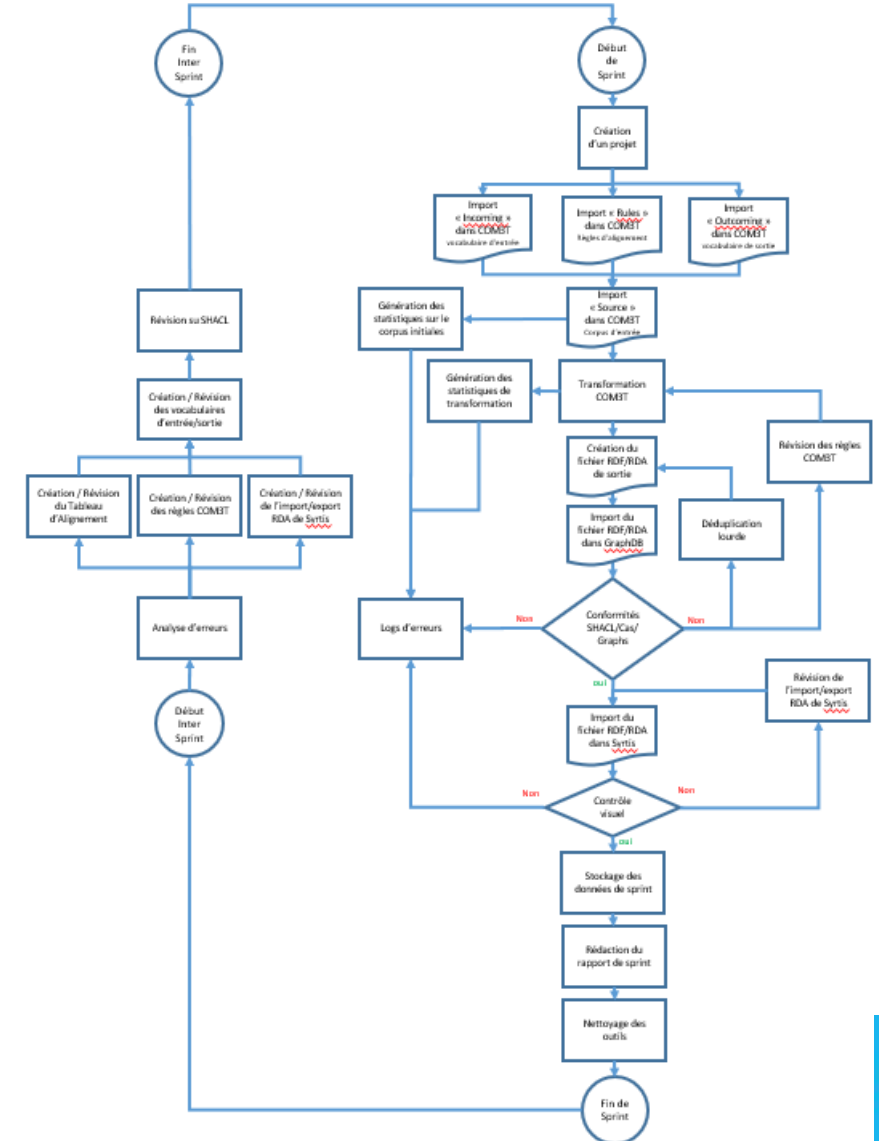
3 main phases :

- Implementation of the interface
- Conversion of the dataset
- Full-scale experimentation after migration

2/ METHODOLOGY AND TOOLS

WORKING TOGETHER ON SPRINTS

- Methodology document : details analysis and validation work for each sprint
- Flowchart : specifies the order and links between each sprint stage
- Tools and guides used on a daily basis: UNIMARC handbook, methodological guides produced by Abes, RDA Toolkit and RDA Registry Official



2/ METHODOLOGY AND TOOLS

14

WORKING TOGETHER ON SPRINTS

Key stages :

- Analysis of the new sprint records (various tools and statistics)
- First migration as the rules set stand / analyze of the results and the statistics
- Correction stage (involves tests and formalization of improvements to be made)
- Second migration : new general migration with improved rules

2/ METHODOLOGY AND TOOLS

15

WORKING TOGETHER ON SPRINTS

Four principal validation tools :

- Examples of ideal RDA records coded by hand in turtle language
- Use GraphDB and its SPARQL endpoint
- Production of a SHACL from our mapping table, usable from GraphDB
- Syrtis visualisation : data in the perspective of a library catalog



*SHACL O'Neal
but the A.I. used
doesn't know
Shaquille O'Neal

3/ FIRST RESULTS AND FUTURE DEVELOPMENTS

3/ FIRST RESULTS AND FUTURE DEVELOPEMENTS

17

SPRINTS 0, 1 AND AFTER

- Testing of the LRM-Factory / implementaton of bibliographic patterns in CoM3T : sprints
- Sprint 0 : 'machine' test
- Sprint 1 : two batches of data of different sizes and creation of the first simple rules in CoM3T
- Sprint 2 : all the steps of the process again, larger number of rules implemented in CoM3T
- Report (internal working document) written for each sprint : comparison of the new sprint with the previous ones.



*We launched sprint_0

3/ FIRST RESULTS AND FUTURE DEVELOPEMENTS

18

SHARING OUR PROJECT

- Contacts and interviews with many people and organizations (committees and authorities, international contacts)
- Opening of a public GitHub “LRM-Factory”
<https://github.com/LRM-Factory>

MERCI !

THANK YOU FOR YOUR ATTENTION !

- carole.bruno@chu-lyon.fr
- marie.bastien@chu-lyon.fr
- morgane.sedoud@chu-lyon.fr

<https://github.com/LRM-Factory>

www.chu-lyon.fr



HCL
HOSPICES CIVILS
DE LYON