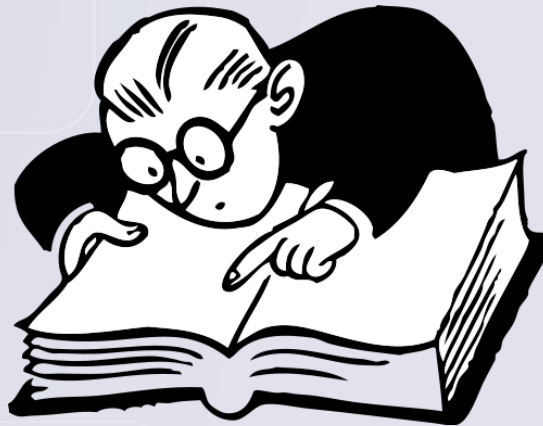


A Survey of FRBRization Techniques

Joffrey Decourselle, Fabien Duchateau, and Nicolas Lumineau

16/09/2015 – TPDL 2015, Poznań, Poland

Too many Online Public Access Catalogs don't meet user needs !



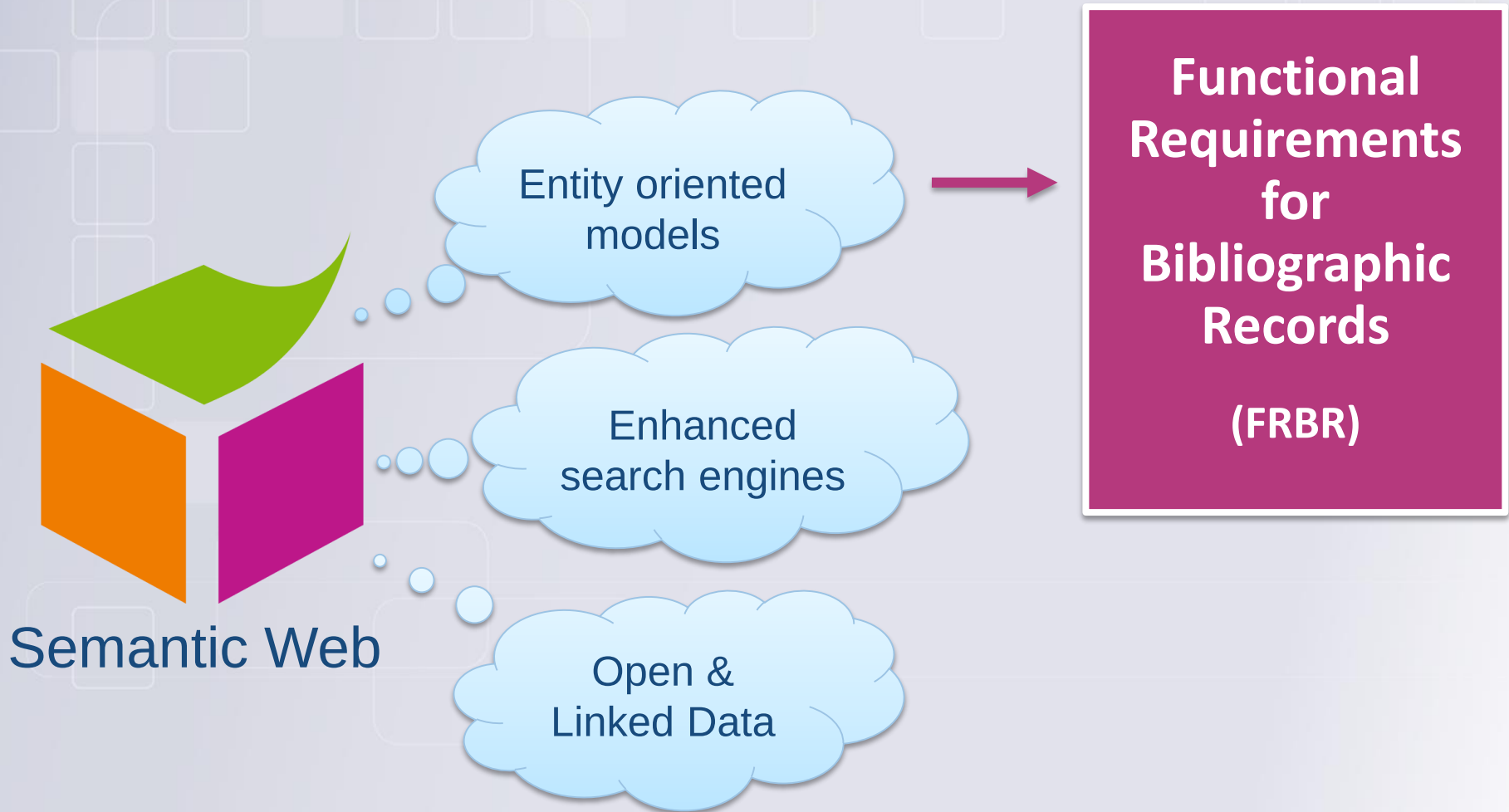
Borgman, C. L. (1996). Why are online catalogs still hard to use?. *JASIS*, 47(7), 493-503.

Yee, M. M. (2005). FRBRization: A method for turning online public finding lists into online public catalogs. *Information technology and libraries*, 24(3).

Today's reality...

- Many catalogs based on flat models (e.g., MARC)
- Solutions with fixed schema
- Very basic search engine
- Bad graphical interfaces

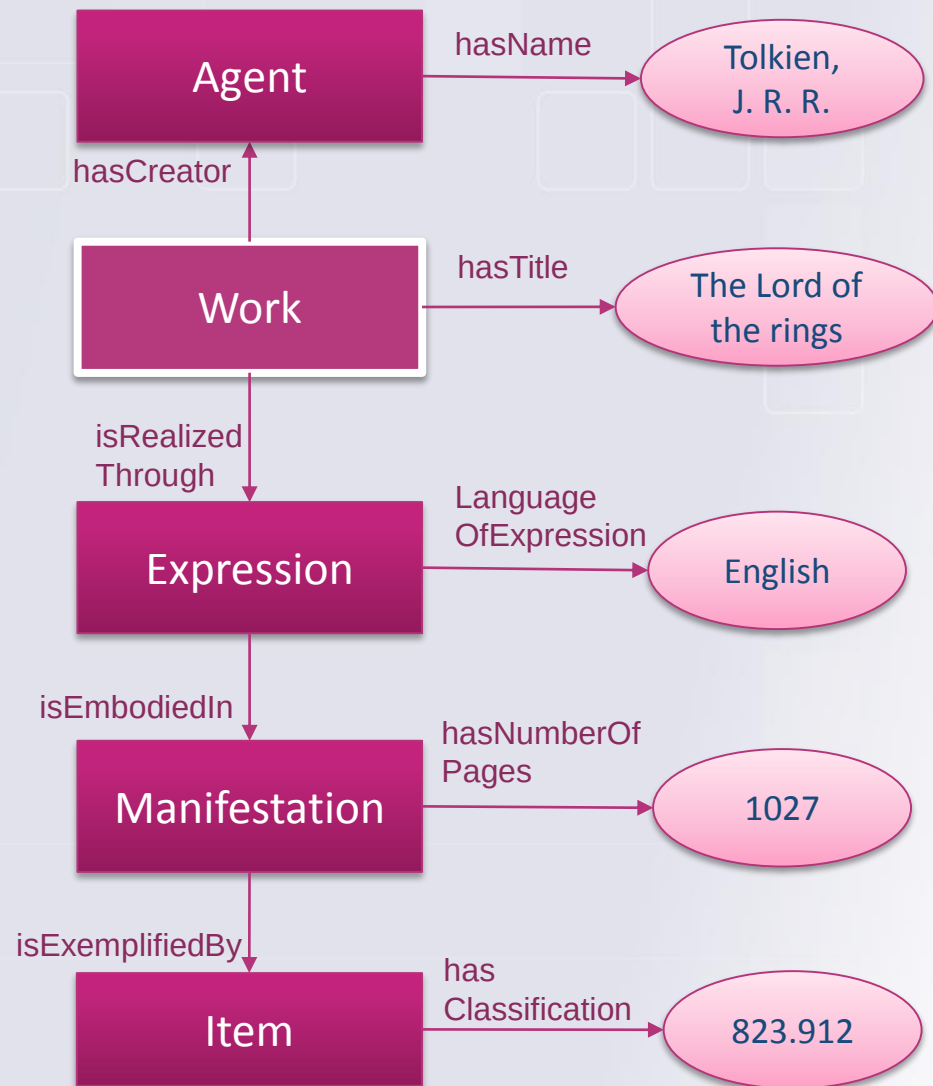
Semantic Web can help improving navigation and enrichment in cultural heritage catalogs



Functional Requirements for Bibliographic Records (FRBR)

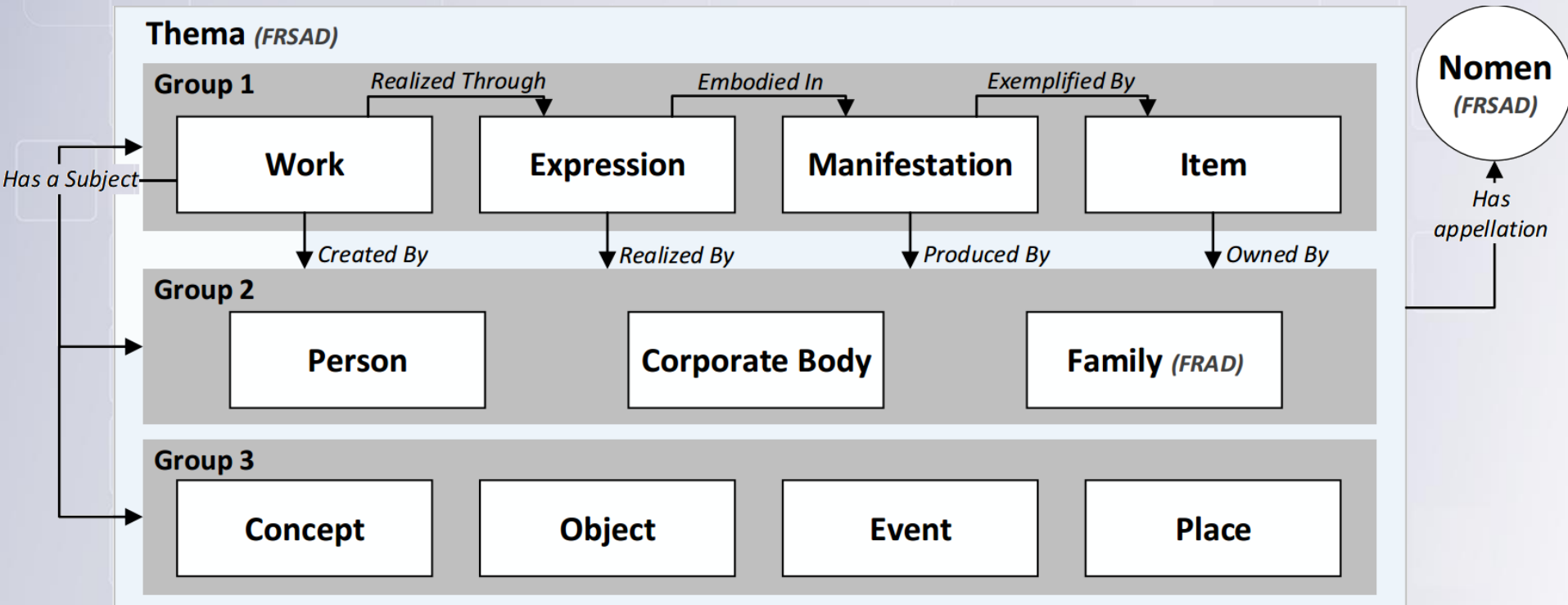
Benefits of FRBR:

- Enhanced representation of cultural content
- Improved searching and visualization features
- Facilities for semantic enrichment



Buchanan, G.: FRBR: enriching and integrating digital libraries. 6th ACM/IEEE-CS Joint Conference on Digital Libraries. 2006

The FRBR conceptual model and main extensions



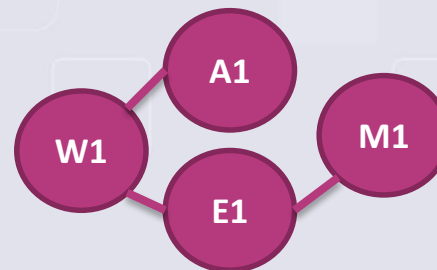
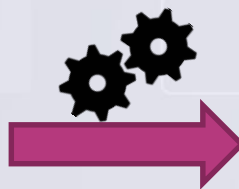
International Federation of Library Associations, & Institutions. Section on Cataloguing. Standing Committee. (1998). *Functional requirements for bibliographic records: final report* (Vol. 19). IFLA Study Group on the Functional Requirements for Bibliographic Records (Ed.). KG Saur Verlag GmbH & Company.

Patton, G. E. (2007). An Introduction to Functional Requirements for Authority Data (FRAD). *Understanding FRBR: what it is and how it will affect our retrieval tools*, 21-27.

FRBRization



Catalog



FRBR

■ Existing surveys about FRBRization techniques:

- **Dickey – 2008:** Benefits of FRBRization

- **Zhang – 2009:** FRBR projects

- **Aalberg – 2013:** Challenges of FRBRization

Dickey, T.J.: FRBRization of a Library Catalog: Better Collocation of Records, Leading to Enhanced Search, Retrieval, and Display. Information Technology & Libraries (2008)

Zhang, Y., Salaba, A.: Implementing FRBR in libraries: key issues and future directions. Neal-Schuman Publishers (2009)

Aalberg, T., Zumer, M.: The value of MARC data, or, challenges of frbrisation. Journal of Documentation (2013)

Limits & Needs

■ Limits of previous surveys

- Mainly focused on specific goal in mind
- Listing of FRBRization solutions at too larger scope

■ Our contributions

- Criteria of FRBRization process
- Classification of existing solutions

Outline

- Introduction
- A novel classification
 - Criteria for FRBRization techniques
 - Classification of existing solutions
 - Observations
- Improving FRBRization process
- Future works

Elements of the classification

- Identification of three criteria
 - Type of FRBRization
 - Model expressiveness
 - Specific enhancements
- Selected FRBRization techniques

Work-Set
Algorithm

Display
Tool

TelPlus

Extensible
Catalog

Variations
FRBR

Primo

LibFRBR

FRBR-ML

Virtua
VTLS

We have excluded projects which have not detailed any process for automating the transformation of a catalog.

1st Criterion : Type of FRBRization

Two major strategies in the literature:

■ Grouping records¹

Generation of descriptive keys for each single record, clustering of records to deduce higher semantic levels according to FRBR

■ Rule-based extraction²

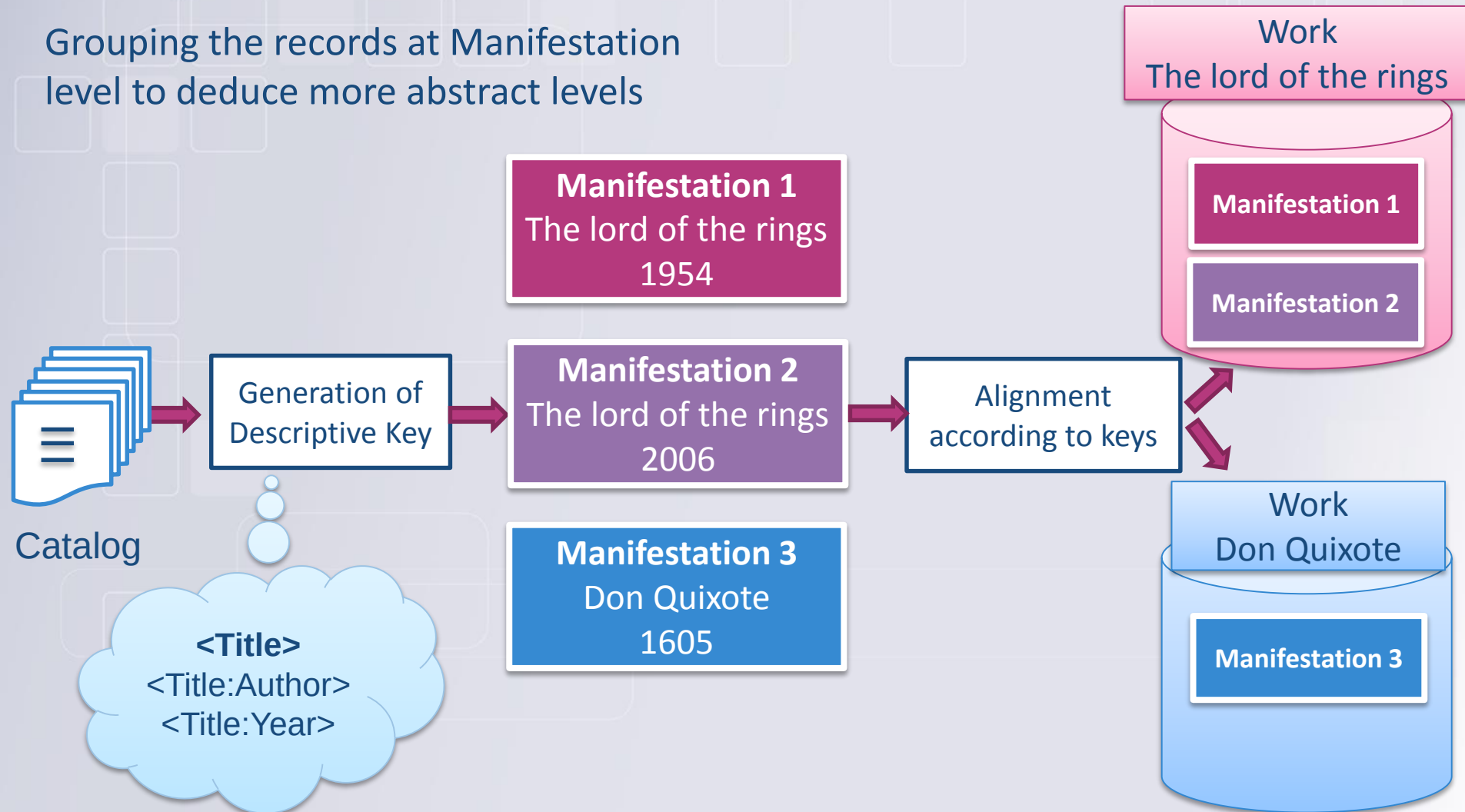
Application of mapping rules on each record to generate entities according to FRBR

1 : Hickey, T.B., O'Neill, E.T., Toves, J.: Experiments with the IFLA functional requirements for bibliographic records (FRBR). D-Lib magazine (2002)

2 : Aalberg, T.: A Process and Tool for the Conversion of MARC Records to a Normalized FRBR Implementation. LNCS: Digital Libraries: Achievements, Challenges and Opportunities (2006)

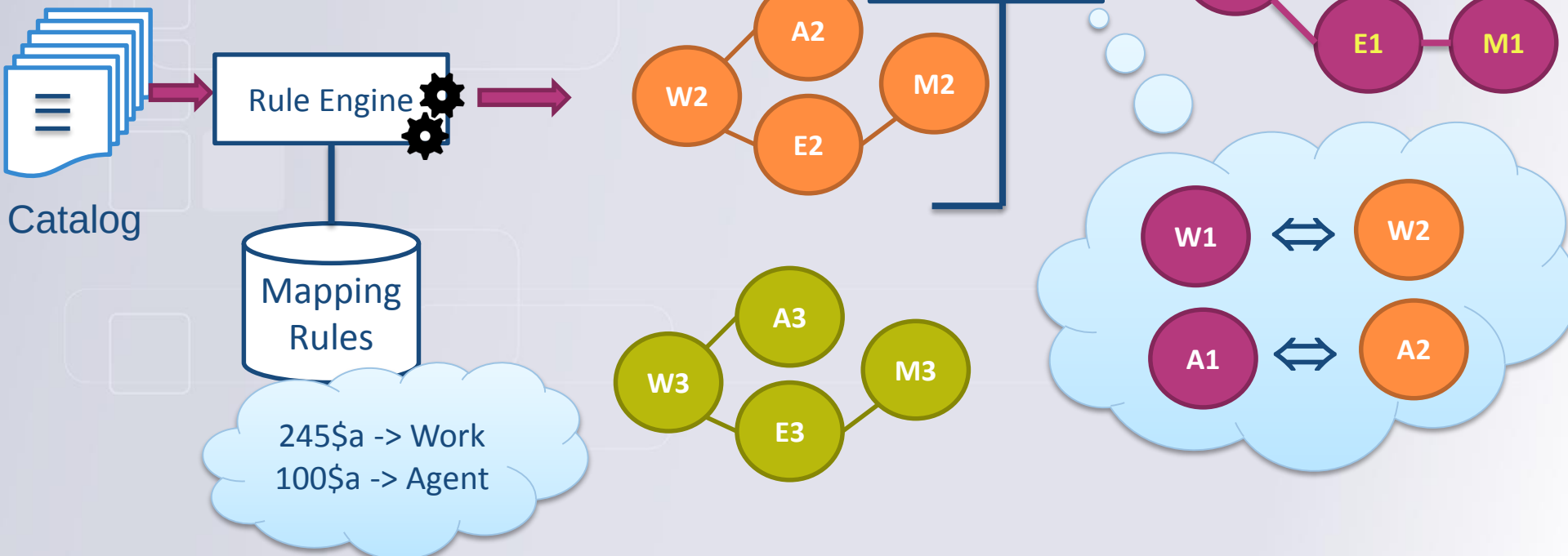
Type of FRBRization: Grouping Records

Grouping the records at Manifestation level to deduce more abstract levels



Type of FRBRization: Rule-based Extraction

Generating a FRBR network for each record and then launch a deduplication step



2nd Criterion: Model Expressiveness

The model expressiveness represents the level of FRBR implementation within models designed to receive FRBRized data

■ Limited Model

1st group of FRBR

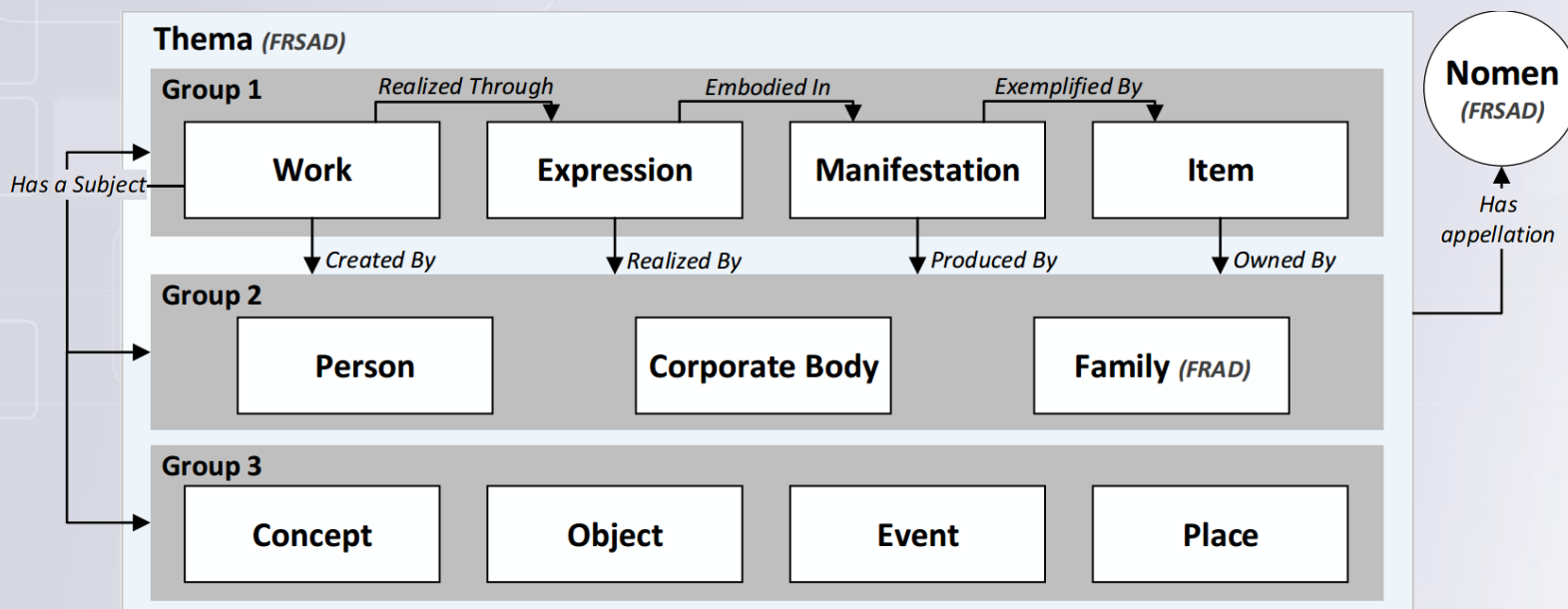
■ Standard Model

3 groups of FRBR

■ Enhanced Model

Standard Model

+ Significant modification



3rd Criterion: Specific Enhancements

All additional steps, improvements of algorithms or interoperability enhancements done on the FRBRization process

Examples:

- **Added Entries Exploitation**

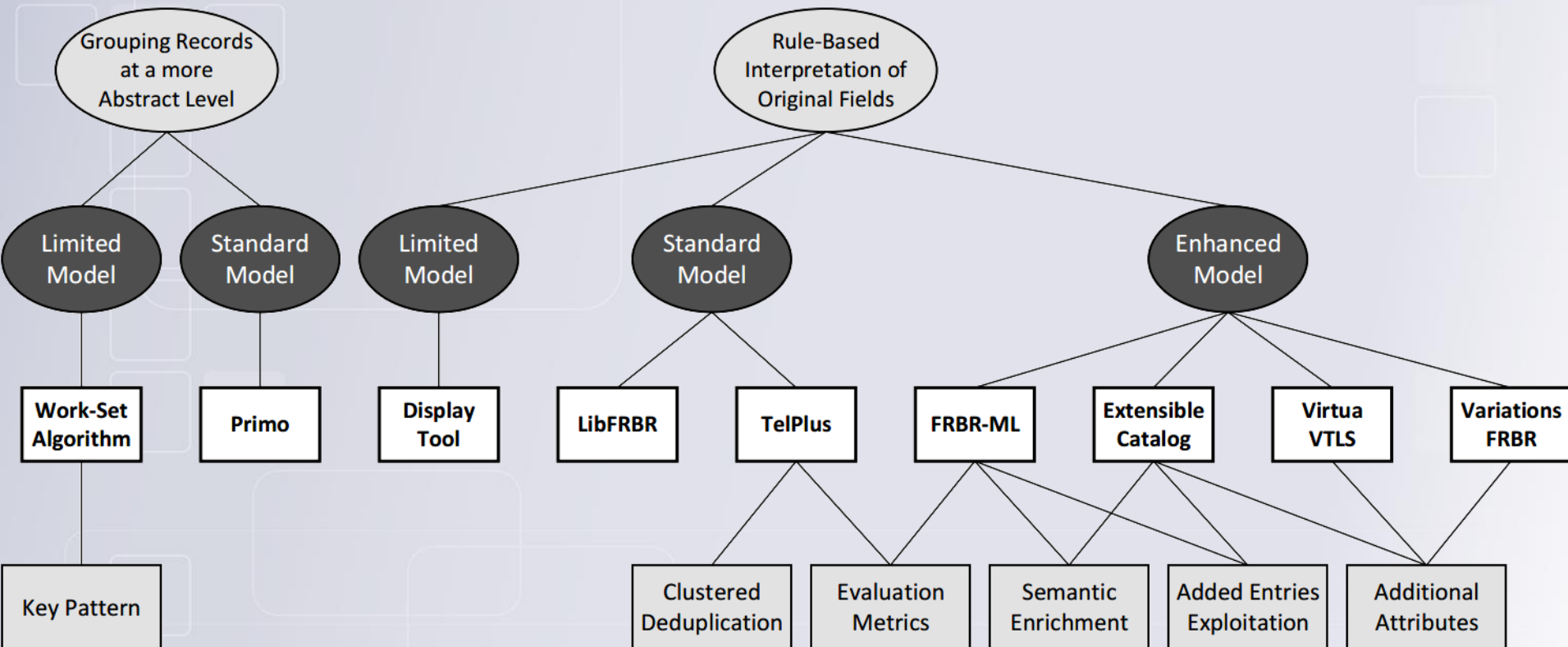
- Using of 700 fields from MARC21

- **Clustered Deduplication**

- Clusters of records for distributed computing

- ...

Classification of FRBRization techniques



Observations

- Both quality and performance enhancements
- Only few propositions for evaluating the process
- Many rules hardcoded (mainly for MARC21)
- No proposition for automatic tuning

Outline

- Introduction
- A novel classification
 - Criteria for FRBRization techniques
 - Classification of existing solutions
 - Observations
- Improving FRBRization process
- Future works

Reality of existing catalogs

```
<datafield tag="200" ind1="|" ind2=" ">  
  <subfield code="a">The Lord Of The Ring = Ringenes herre</subfield>  
  <subfield code="a">The two towers</subfield>  
  <subfield code="f">J.R.R. Tolkien</subfield>  
  <subfield code="g">trans. by Eilev Groven</subfield>  
</datafield>
```

Metadata
Inconsistencies

```
<datafield tag="200" ind1="|" ind2=" ">  
  <subfield code="a">The Lord Of The Ring</subfield>  
  <subfield code="a">The two towers</subfield>  
</datafield>
```

Duplicate
subfields

```
<datafield tag="100" ind1=" " ind2=" ">  
  <subfield code="a">19900801d1989      y0frey0103  ba</subfield>  
</datafield>
```

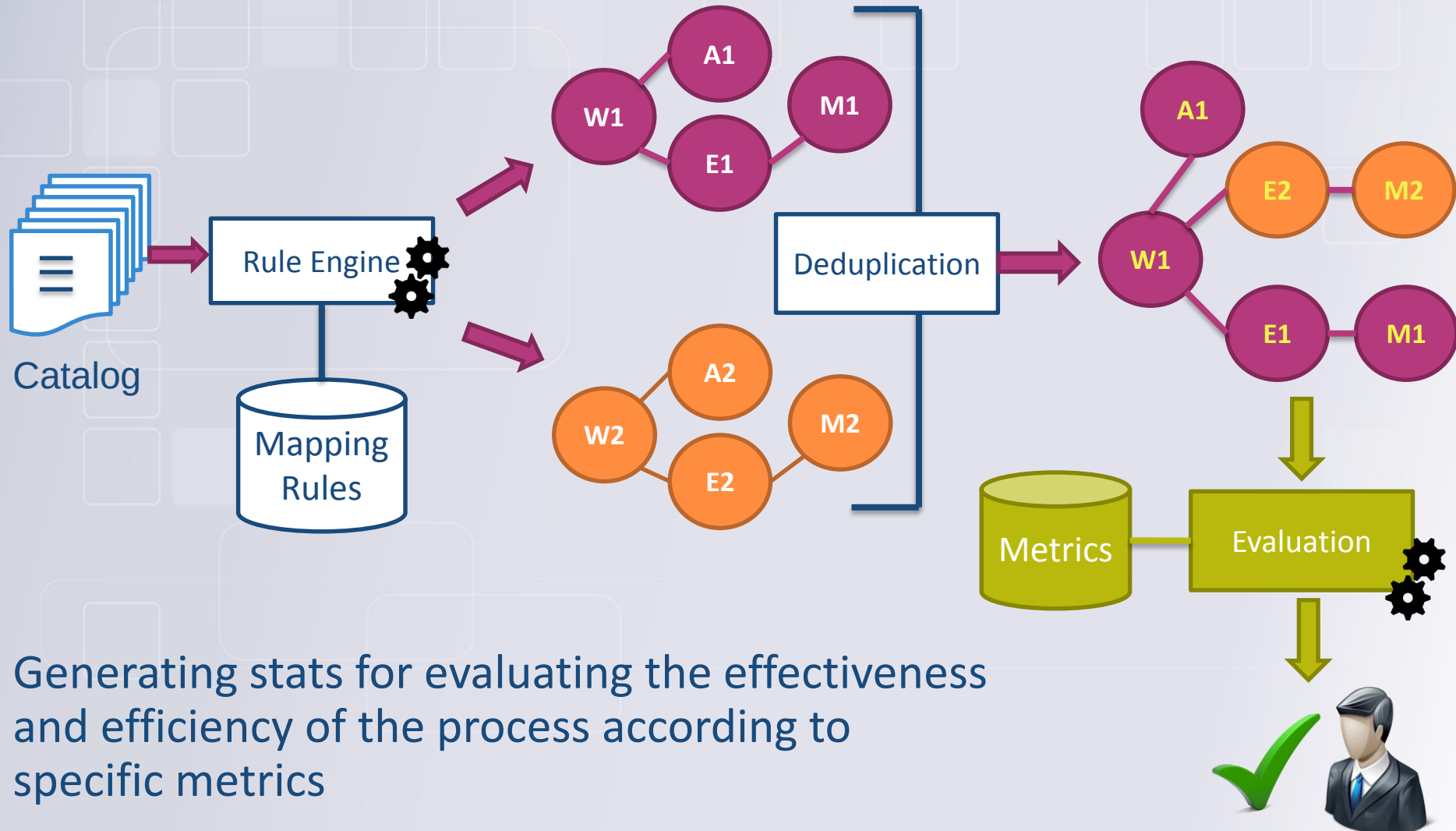
Fields where the
position of
characters matters

Evaluating the process

The expert must be able to answer these questions:

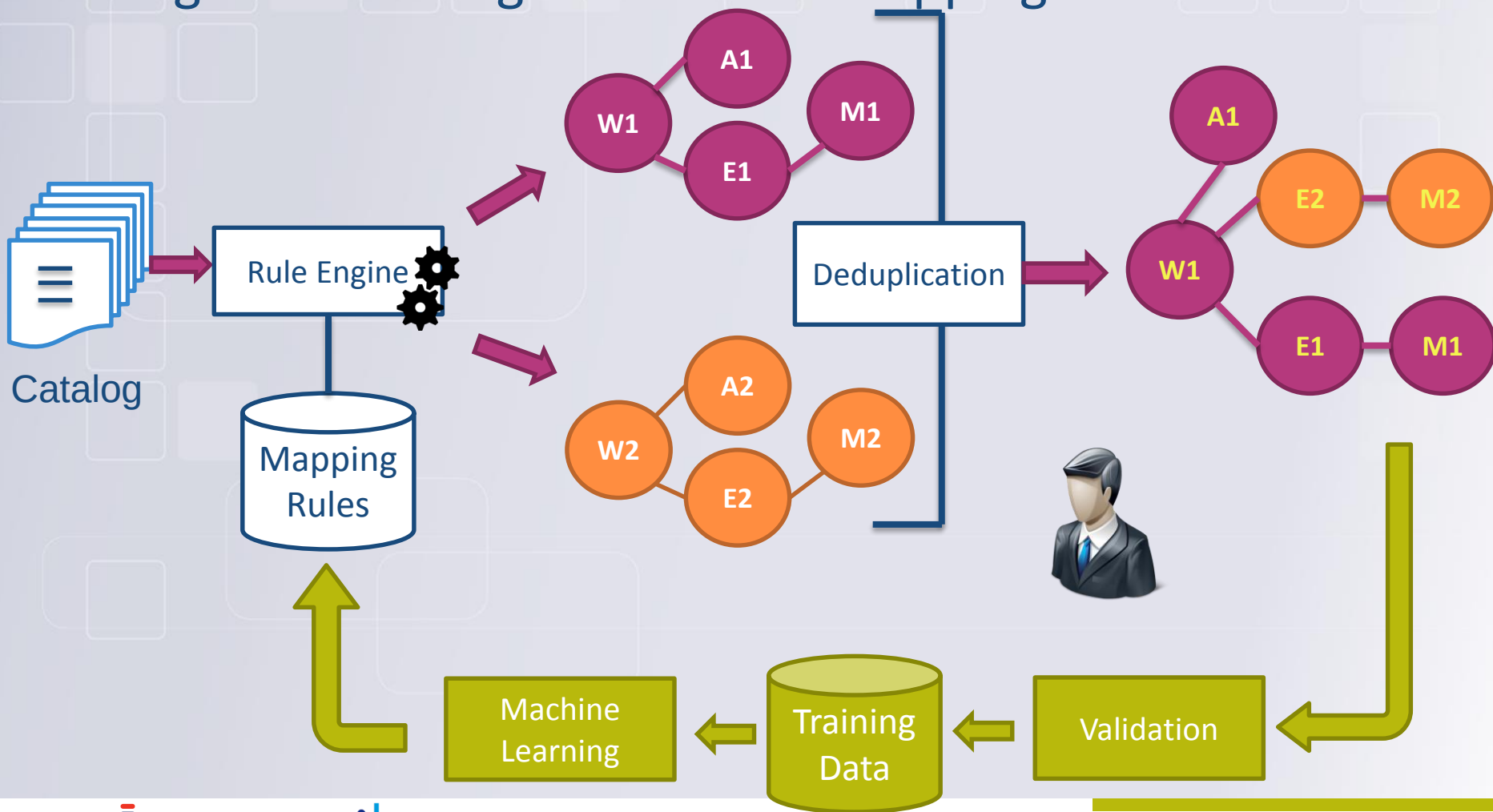
- Is there any loss of data during the process ?
- What is the total time (tuning, process & validation) for processing a record ?

Need for new evaluation metrics



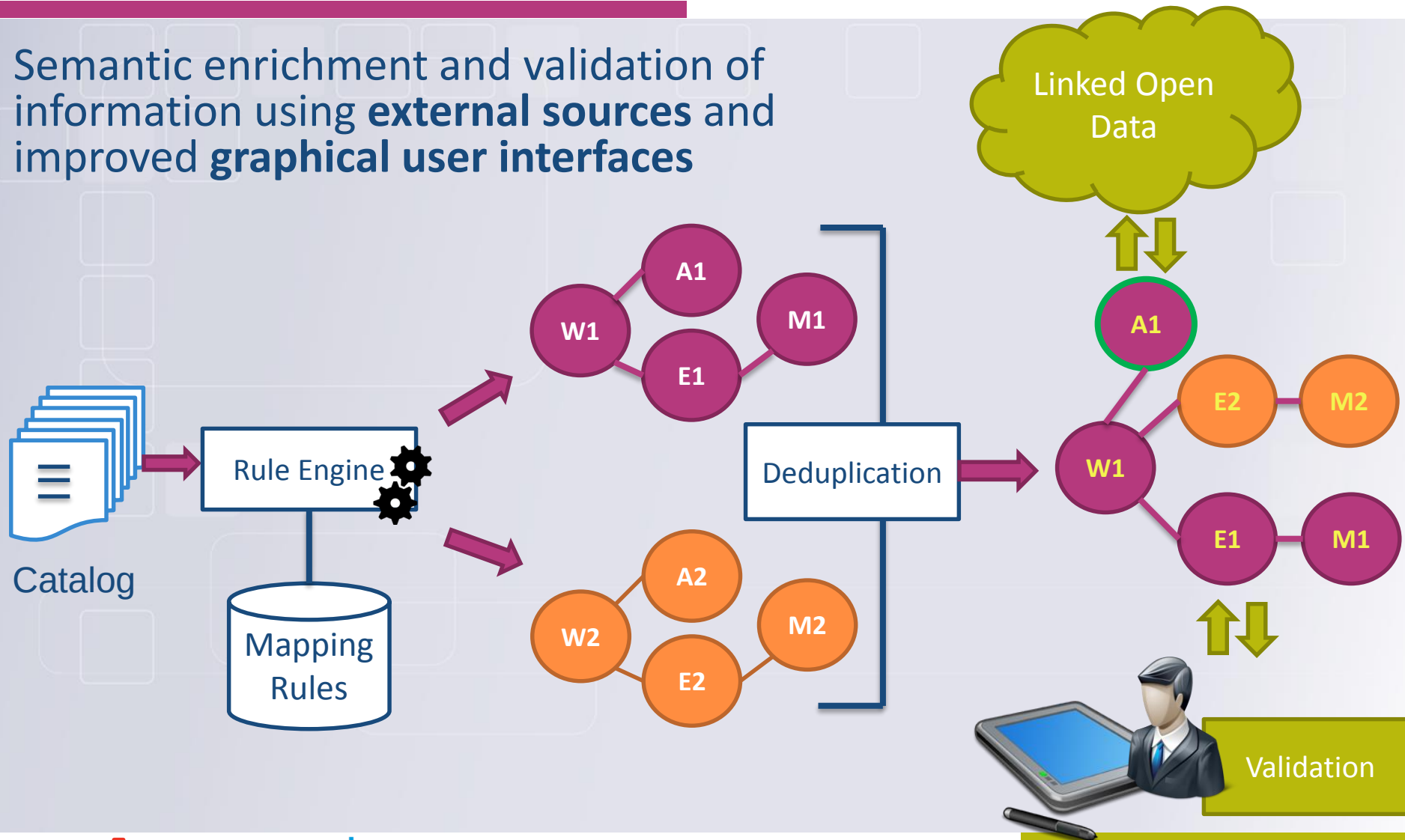
Need for (semi-)automated tuning

Enabling automated generation of mapping rules



Need for assisted validation

Semantic enrichment and validation of information using **external sources** and improved **graphical user interfaces**



Outline

- Introduction
- A novel classification
 - Criteria for FRBRization techniques
 - Classification of existing solutions
 - Observations
- Improving FRBRization process
- Future work

Future work

- Based on our survey:
 - Building a rule-based application for FRBRization including all specific enhancements, mainly:
 - Semi-automated tuning
 - Semantic enrichment with external sources (e.g., LOD)
 - Benchmark of this application with existing solutions according to new metrics and datasets

Thank you !

To get more details about our projects:



<http://liris.cnrs.fr/diricks/>



syrtis
by progilone

<http://www.progilone.fr/en/syrtis>