Appendix: BIB-R: a Benchmark for the Interpretation of Bibliographic Records

Joffrey Decourselle¹, Fabien Duchateau¹, Trond Aalberg², Naimdjon Takhirov³, and Nicolas Lumineau¹

¹ LIRIS, UMR5205, Université Claude Bernard Lyon 1, Lyon, France
firstname.lastname@liris.cnrs.fr
² NTNU, Trondheim, Norway trondaal@idi.ntnu.no
³ Westerdals - Oslo School of Arts, Communication and Technology - Faculty of Technology, Oslo, Norway taknai@westerdals.no

Abstract. This appendix contains all experiment results for the submission to TPDL2016. In the cultural heritage domain, cataloging items is a crucial task, which has been performed in decades based on the MAchine Readable Cataloguing (MARC) format. In a context which promotes the use of semantics and sharing of information, MARC has demonstrated its limitations. The Functional Requirements for Bibliographic Records (FRBR), expected to be a successor of MARC, is a semantic model for representing cultural items. The complex transformation of MARC catalogs to FRBR catalogs (FRBRization) led to the proposition of various tools and approaches. However these projects and the results they achieve are difficult to compare on a fair basis due to a lack of common datasets and appropriate metrics. Our contributions fill this gap by proposing the first public benchmark for the FRBRization process.

Keywords: benchmark, migration, record interpretation, FRBRization, LRM, FRBR, MARC, dataset, evaluation metric

1 Formal notation of pre-FRBRization metrics

Table 1 provides formal notations for the predictive metrics. Each metric computes a percentage defined as the number of records concerned with the given pattern/issue divided by the total number of records. The detection of these issues uses the first-order logic notation for genericity reasons. In the table, we define a record $r$ which belongs to a set of records $R$. We note $r \rightarrow \text{`specification'}$ the fact that the record $r$ implements a given specification such as catalog rule (e.g., AACR2) or punctuation (e.g., ISBD). This record is composed of a set of fields, each of them representing a concept. For instance, the notation $f \in r, f \rightarrow \text{`title'}$ indicates that the field $f$ belongs to the record $r$ and it stands for a title in this context. We also introduce a rule $\ell$ from a set $\mathcal{L}$. A rule contains a condition denoted $\text{cond}_\ell$, which satisfies a field $f$ when $\text{cond}_\ell \models f$. Some fields may encode specific values that needs to be extracted. We define $\varphi(\text{value})$ as a function which extracts frequent patterns from a given value by analyzing the whole collection. It returns an empty set if no pattern is detected, or a list of pattern otherwise (e.g., "illustrated by"). Various approaches enable this extraction of patterns, traditionally based on machine learning techniques [2]. To illustrate the table, let us see the metric MID, which stands for the percentage of records in which the record identifier is missing. Its formal notation means that whatever the field $f$ in a record $r$, this field $f$ does not correspond to the concept 'record identifier'. Since this is a generic definition, it can be instantiated for a specific format. For example, the UNIMARC notation for the MID metric could be written: $\forall f \in r, f \neq '001'$. In MARC21, the metric MPD about missing publication date can be redefined as: $\forall f \in r, f \neq '260\$c'$. A more complex notation deals with the metric CPN, which consists of measuring the percentage of records that implements a specification (catalog rule or punctuation), or which contains a field $f$ used to store local data (traditionally the 900 fields in MARC21) or a field with recurrent values (e.g., the Norwegian...
country code could be written "NO" or "NOR"). Finally, bibliographic patterns can be detected using multiple fields. Note that when we indicate value\(f \in \{'Derivations'\}\), it means that the value of \(f\) corresponds to a Derivation (e.g., \(trl\) for translator in MARC21\(^4\)) according to Tillet and Riva’s taxonomies [1]. The metric MR can be decomposed into more detailed metrics such as MR-AUG to calculate the percentage of missing rules for detecting all augmentations. In the same fashion, we define the metrics MR-DER for derivations, MR-AGG for aggregations, MR-COW for complementary works, and MR-CPN for cataloging practices. Their formal notation is obtained by combining the definitions of both MR and the given pattern/issue.

2 Formal notation of post-FRBRization metrics

Table 2 provides first-order logic notations for the detection of post-FRBRization issues. During evaluation, we compare two collections, \(\mathcal{T}\) which is produced by a tool and the expert collection \(\mathcal{E}\). This comparison depends on the type of data. Consider the data \(e \in \mathcal{E}\) and \(t \in \mathcal{T}\). When dealing with entities, the type of entity and the value of its main label (e.g., title, name) needs to be verified:

\[
e \equiv t \iff \text{type}_e = \text{type}_t \land \text{value}_e = \text{value}_t
\]

For relationships, the checking is performed based on the type of relationship and the two linked entities:

\[
e \equiv t \iff \text{type}_e = \text{type}_t \land \text{entity}_{1e} = \text{entity}_{1t} \land \text{entity}_{2e} = \text{entity}_{2t}
\]

Finally, the properties are compared according to their type, their owner (entity) and their value:

\[
e \equiv t \iff \text{type}_e = \text{type}_t \land \text{entity}_e = \text{entity}_t \land \text{value}_e = \text{value}_t
\]

Besides, it is not possible to verify information about bibliographic patterns without annotation in the expert collection. Thus we define the set \(\mathcal{E}' \in \mathcal{E}\) which includes all main elements of a pattern (i.e., the main entity and the main relationship) and the set \(\mathcal{E}'' \in \mathcal{E}\) which contains all secondary elements of a pattern.

Let us describe Table 2. First, the metric MD is related to the missing data issue. The formal notation states that such data \(e\) appears in the expert collection \(\mathcal{E}\) but has no equivalence in the tool’s collection \(\mathcal{T}\). This metric computes the ratio between the number of missing data and the total number of data in the expert collection. It can be redefined for each type of data, i.e., MD-E for entities, MD-R for relationships and MD-P for properties. The metric IAD deals with incorrectly added data, i.e., which appear in \(\mathcal{T}\) but not in \(\mathcal{E}\). It is defined as the number of incorrect data in \(\mathcal{T}\) divided by the total number of data in \(\mathcal{T}\). Similarly to MD, the metric IAD can be redefined according to the data type. The metric DLE relates to errors in external link, i.e., either the link does not exist in \(\mathcal{E}\) or it has a different value for the same external source. The metric calculates precision, i.e., the number of erroneous links in \(\mathcal{T}\) divided by the total number of links in \(\mathcal{T}\). The metric SMD computes the amount of data with a different semantics (usually a subsumption noted \(t \subset e\)) with regards to the total amount of data. The metric MEND measures the percentage of main entities in \(\mathcal{T}\) that have been correctly detected among all main entities in \(\mathcal{E}'\). To check the correctness of a detection, a main entity \(e'\) should have an equivalent entity \(t\) in the tool’s collection. Note that the metrics MRND and ESE have a similar definition due to our generic notation.

\(^4\) MARC21 code list for relators
<table>
<thead>
<tr>
<th>Related pattern/issue</th>
<th>Formal notation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AUG</strong> Augmentation</td>
<td>$f \in r, (f \leadsto \text{‘secondary resp.’}) \lor (f \leadsto \text{‘title’}) \land \varphi(value_f) \neq \emptyset \lor (f \leadsto \text{‘relator code’}) \land value_f \in {\text{‘Augmentations’}}$</td>
</tr>
<tr>
<td><strong>DER</strong> Derivation</td>
<td>$f \in r, (f \leadsto \text{‘original language’} \land \exists f_2 \in r \land (f \leadsto \text{‘resource language’}) \lor (f \leadsto \text{‘title’}) \land \varphi(value_f) \neq \emptyset \lor (f \leadsto \text{‘relator code’}) \land value_f \in {\text{‘Derivations’}} \lor (f \leadsto \text{‘variant title’})$</td>
</tr>
<tr>
<td><strong>AGG</strong> Aggregation</td>
<td>$f \in r, (f \leadsto \text{‘collection/ensemble link’}) \lor (f \leadsto \text{‘collective title’}) \lor (f \leadsto \text{‘proper title’}) \land \exists f^1, \ldots, f^k \in r \land (f \leadsto \text{‘proper title’}) \land \ldots \land (f \leadsto \text{‘variant title’})$</td>
</tr>
<tr>
<td><strong>COW</strong> Complementary works</td>
<td>$f \in r, (f \leadsto \text{‘title’}) \land \exists f_2 (f_2 \leadsto \text{‘subtitle’}) \land value_{f_2} \neq value_f) \lor (f \leadsto \text{‘title’}) \land \varphi(value_f) \neq \emptyset \lor (f \leadsto \text{‘linking field’}) \lor (f \leadsto \text{‘note’}) \lor (f \leadsto \text{‘added entry’})$</td>
</tr>
<tr>
<td><strong>MID</strong> Missing record identifier</td>
<td>$\forall f \in r, f \not\sim \text{‘record identifier’}$</td>
</tr>
<tr>
<td><strong>MPD</strong> Missing publication date</td>
<td>$\forall f \in r, f \not\sim \text{‘publication date’}$</td>
</tr>
<tr>
<td><strong>MTF</strong> Missing type and form</td>
<td>$\forall f \in r, f \not\sim \text{‘type/form’}$</td>
</tr>
<tr>
<td><strong>TLE</strong> Title linkage error</td>
<td>$f \in r, f \leadsto \text{‘auth. title’} \land \exists r_2 \in R(f_2 \in r_2 \land f_2 \leadsto \text{‘record identifier’}) \land value_{f_2} = value_f$</td>
</tr>
<tr>
<td><strong>MOT</strong> Missing uniform title</td>
<td>$\forall f \in r, f \not\sim \text{‘uniform title’}$</td>
</tr>
<tr>
<td><strong>RLE</strong> Resp. linking error</td>
<td>$f \in r, f \leadsto \text{‘auth. resp.’} \land \exists r_2 \in R(f_2 \in r_2 \land f_2 \leadsto \text{‘record identifier’}) \land value_{f_2} = value_f$</td>
</tr>
<tr>
<td><strong>MRC</strong> Missing relator code</td>
<td>$\forall f \in r, f \not\sim \text{‘relator code’}$</td>
</tr>
<tr>
<td><strong>MAR</strong> Missing auth. resp.</td>
<td>$\forall f \in r, f \not\sim \text{‘auth. resp.’}$</td>
</tr>
<tr>
<td><strong>CPN</strong> Cataloging practices and norms</td>
<td>$r \leadsto \text{‘catalog rule’} \lor r \leadsto \text{‘punctuation’} \lor (f \in r, f \leadsto \text{‘local data’} \lor \varphi(value_f) \neq \emptyset)$</td>
</tr>
<tr>
<td><strong>MR</strong> Missing rule</td>
<td>$f \in r, \forall \ell \in \mathcal{L}, \text{cond}_\ell \not\sim f$</td>
</tr>
<tr>
<td><strong>UR</strong> Useless rule</td>
<td>$\ell \in \mathcal{L}, \forall r \in \mathcal{R}, \forall f \in r, \text{cond}_\ell \not\sim f$</td>
</tr>
<tr>
<td><strong>CR</strong> Conflicting rules</td>
<td>$\forall \ell_1 \in \mathcal{L}, \forall \ell_2 \in \mathcal{L}, \ell_1 \neq \ell_2 \land \text{cond}<em>{\ell_1} \equiv \text{cond}</em>{\ell_2}$</td>
</tr>
</tbody>
</table>

Table 1. List of pre-FRBRization metrics
<table>
<thead>
<tr>
<th>Related issue</th>
<th>Formal notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD</td>
<td>Missing data</td>
</tr>
<tr>
<td>IAD</td>
<td>Incorrectly added data</td>
</tr>
<tr>
<td>DLE</td>
<td>Data linking error</td>
</tr>
<tr>
<td>SMD</td>
<td>Semantic mismatch data</td>
</tr>
<tr>
<td>MEND</td>
<td>Main entity not detected</td>
</tr>
<tr>
<td>MRND</td>
<td>Main relationship not detected</td>
</tr>
<tr>
<td>ESE</td>
<td>Error(s) in secondary elements</td>
</tr>
</tbody>
</table>

*Table 2.* List of post-FRBRization metrics
3 Assessing strengths and weaknesses - results by test

This first experiment aims at demonstrating the benefit of the dataset T42 when it comes to evaluating the strengths and weaknesses of FRBRization tools. For the three tools, we have run each test from the dataset T42 and the evaluation is performed using post-FRBRization metrics. In other words, each tool has produced a FRBR collection for each test, and these generated FRBR collections have been compared to the expert ones provided in the benchmark. A basic set of rules is available with each tool. For equity reasons, we have not tuned the tools by updating their set of rules.

In the following plots, the results are organized by tool and by test.

Fig. 1. Quality results for FRBR-ML with dataset T42 tests 1.x

Fig. 2. Quality results for FRBR-ML with dataset T42 tests 2.x
Fig. 3. Quality results for FRBR-ML with dataset T42 tests 3.x

Fig. 4. Quality results for FRBR-ML with dataset T42 tests 4.x

Fig. 5. Quality results for FRBR-ML with dataset T42 tests 5.x
Fig. 6. Quality results for Variation/VFRBR with dataset T42 tests 1.x

Fig. 7. Quality results for Variation/VFRBR with dataset T42 tests 2.x
Fig. 8. Quality results for Variation/VFRBR with dataset T42 tests 3.x

Fig. 9. Quality results for Variation/VFRBR with dataset T42 tests 4.x

Fig. 10. Quality results for Variation/VFRBR with dataset T42 tests 5.x
Fig. 11. Quality results for Extensible Catalog with dataset T42 tests 1.x

Fig. 12. Quality results for Extensible Catalog with dataset T42 tests 2.x
Fig. 13. Quality results for Extensible Catalog with dataset T42 tests 3.x

Fig. 14. Quality results for Extensible Catalog with dataset T42 tests 4.x

Fig. 15. Quality results for Extensible Catalog with dataset T42 tests 5.x
4 Assessing strengths and weaknesses - results by metric

This first experiment aims at demonstrating the benefit of the dataset T42 when it comes to evaluating the strengths and weaknesses of FRBRization tools. For the three tools, we have run each test from the dataset T42 and the evaluation is performed using post-FRBRization metrics. In other words, each tool has produced a FRBR collection for each test, and these generated FRBR collections have been compared to the expert ones provided in the benchmark. A basic set of rules is available with each tool. For equity reasons, we have not tuned the tools by updating their set of rules.

In the following plots, the results are organized by tool and by metric.

Fig. 16. Quality results for FRBR-ML with dataset T42 and metric MEND
Fig. 17. Quality results for FRBR-ML with dataset T42 and metric MRND

Fig. 18. Quality results for FRBR-ML with dataset T42 and metric ESE
Fig. 19. Quality results for FRBR-ML with dataset T42 and metric MD

Fig. 20. Quality results for FRBR-ML with dataset T42 and metric MD-E
Fig. 21. Quality results for FRBR-ML with dataset T42 and metric MD-R.

Fig. 22. Quality results for FRBR-ML with dataset T42 and metric MD-P.
**Fig. 23.** Quality results for FRBR-ML with dataset T42 and metric IAD

**Fig. 24.** Quality results for FRBR-ML with dataset T42 and metric IAD-E
Fig. 25. Quality results for FRBR-ML with dataset T42 and metric IAD-R

Fig. 26. Quality results for FRBR-ML with dataset T42 and metric IAD-P
Fig. 27. Quality results for FRBR-ML with dataset T42 and metric SMD

Fig. 28. Quality results for FRBR-ML with dataset T42 and metric SMD-E
Fig. 29. Quality results for FRBR-ML with dataset T42 and metric SMD-R

Fig. 30. Quality results for FRBR-ML with dataset T42 and metric SMD-P
Fig. 31. Quality results for Variation/VFRBR with dataset T42 and metric MEND
Fig. 32. Quality results for Variation/VFRBR with dataset T42 and metric MRND

Fig. 33. Quality results for Variation/VFRBR with dataset T42 and metric ESE
Fig. 34. Quality results for Variation/VFRBR with dataset T42 and metric MD

Fig. 35. Quality results for Variation/VFRBR with dataset T42 and metric MD-E
Fig. 36. Quality results for Variation/VFRBR with dataset T42 and metric MD-R

Fig. 37. Quality results for Variation/VFRBR with dataset T42 and metric MD-P
Fig. 38. Quality results for Variation/VFRBR with dataset T42 and metric IAD

Fig. 39. Quality results for Variation/VFRBR with dataset T42 and metric IAD-E
Fig. 40. Quality results for Variation/VFRBR with dataset T42 and metric IAD-R

Fig. 41. Quality results for Variation/VFRBR with dataset T42 and metric IAD-P
Fig. 42. Quality results for Variation/VFRBR with dataset T42 and metric SMD

Fig. 43. Quality results for Variation/VFRBR with dataset T42 and metric SMD-E
Fig. 44. Quality results for Variation/VFRBR with dataset T42 and metric SMD-R.

Fig. 45. Quality results for Variation/VFRBR with dataset T42 and metric SMD-P.
Fig. 46. Quality results for XC with dataset T42 and metric MEND
Fig. 47. Quality results for XC with dataset T42 and metric MRND

Fig. 48. Quality results for XC with dataset T42 and metric ESE
Fig. 49. Quality results for XC with dataset T42 and metric MD

Fig. 50. Quality results for XC with dataset T42 and metric MD-E
Fig. 51. Quality results for XC with dataset T42 and metric MD-R

Fig. 52. Quality results for XC with dataset T42 and metric MD-P
Fig. 53. Quality results for XC with dataset T42 and metric IAD

Fig. 54. Quality results for XC with dataset T42 and metric IAD-E
Fig. 55. Quality results for XC with dataset T42 and metric IAD-R

Fig. 56. Quality results for XC with dataset T42 and metric IAD-P
Fig. 57. Quality results for XC with dataset T42 and metric SMD

Fig. 58. Quality results for XC with dataset T42 and metric SMD-E
Fig. 59. Quality results for XC with dataset T42 and metric SMD-R

Fig. 60. Quality results for XC with dataset and metric SMD-P
5 Comparing tools in real-world context

The objective of this second experiment is to compare FRBRization tools in a real-world context using both FRBRization and post-FRBRization metrics. All tools rely on their basic set of rules (no tuning). Contrary to the dataset T42, all records from the dataset BIB-RCAT come from various institutions. This means that these real-world records can be rather simple or may contain several bibliographic patterns and issues. Note that the FRBRization metrics ETC, ETD and NRT could not be presented, because the tools did not implement these features. Instead, we provide the overall execution time for FRBRizing the dataset BIB-RCAT. The post-FRBRization metric DLE is also not given, since there exists many authority files or knowledge bases (e.g., Linked Open Data) and the expert FRBRized collection cannot include a link for each of these sources.

In the following plots, the results are organized by tool.

![Quality results for FRBR-ML basic (no tuning) on dataset BIB-RCAT](image)

**Fig. 61.** Quality results for FRBR-ML basic (no tuning) on dataset BIB-RCAT
Fig. 62. Quality results for VFRBR on dataset BIB-RCAT

Fig. 63. Quality results for XC on dataset BIB-RCAT
Fig. 64. Quality results for FRBR-ML tuned (enhanced set of rules) on dataset BIB-RCAT
6 Additional information about experiments

In this section, we provide screenshots of the tools that have been used during experiments.

Figure 65 depicts (an extract of) the FRBRized collection of the BIB-RCAT dataset using FRBR-ML.
Mireille l'abeille (Krings, Antoon)
Mireille l'abeille
Antoon Krings / Mireille l'abeille
Antoon Krings / Mireille l'abeille
Antoon Krings / Mireille l'abeille
Antoon Krings / Mireille l'abeille

Un sac de billes (Joffo, Joseph)
Un sac de billes
Joseph Joffo; illustrations Claude Lapointe / Un sac de billes
Joseph Joffo; illustrations Claude Lapointe / Un sac de billes

La Vénus d'Ille (Mérimée, Prosper)
La Vénus d'Ille ; La chambre bleue La Vénus d'Ille
Prosper Mérimée; Prosper Mérimée / La Vénus d'Ille ; La chambre bleue
Mérimée; édition Dominique Fleur-Schulthess, Claudine Zenou-Grinstein / La Vénus d'Ille

Les allumettes suédoises (Sabatier, Robert)
Les allumettes suédoises
Robert Sabatier, lu par Jean Barrier / Les allumettes suédoises
Robert Sabatier / Les allumettes suédoises
Robert Sabatier; illustrations Louis Constantin / Les allumettes suédoises
Robert Sabatier; illustrations Louis Constantin / Les allumettes suédoises
Robert Sabatier; illustrations Louis Constantin / Les allumettes suédoises

Tous à poil ! (Franck, Claire)
Tous à poil !
Claire Franck, Marc Daniau / Tous à poil !

Les contes bleus du chat perché (Aymé, Marcel)
Les contes bleus du chat perché
Marcel Aymé; illustrations Claudine et Roland Sabatier / Les contes bleus du chat perché

Un sac de billes (Bouton, Alain)
Un sac de billes
d'après Joseph Joffo; scénario Alain Bouton; dessins Marc Malès / Un sac de billes

365 contes pour tous les âges (Bloch, Muriel)
365 contes pour tous les âges
Muriel Bloch; illustrations Mireille Vautier / 365 contes pour tous les âges
Muriel Bloch; illustrations Grégoire Solotareff / 365 contes pour tous les âges
Muriel Bloch / 365 contes pour tous les âges

La belle lisse poire du prince de Motordu ()
La belle lisse poire du prince de Motordu ; et 4 autres histoires La belle lisse poire du

Fig. 65. Extract of the FRBRized BIB-RCAT dataset in FRBR-ML
Figure 66 depicts (an extract of) the relationships between FRBR entities in the FRBRized collection of the BIB-RCAT dataset using FRBR-ML.

Fig. 66. Extract of a graph-based visualization of the FRBRized BIB-RCAT dataset in FRBR-ML.
Figure 67 depicts (an extract of) the FRBRized collection of the BIB-RCAT dataset using Variations (Scherzo).
### Works: 207 results for "**:**" as Keyword

1. La Fontaine, Jean de
   "..."
   People: La Fontaine, Jean de
   Copies: See Catalog

2. Brunhoff, Jean de
   "..."
   People: Brunhoff, Jean de
   Copies: See Catalog

3. Guillo, Ren
   "..."
   People: Guillo, Ren
   Copies: See Catalog

4. Daudet, Alphonse
   "..."
   People: Daudet, Alphonse
   Copies: See Catalog

5. Molière
   "..."
   People: Molière
   Copies: See Catalog

### Recordings/Scores: 560 results for "**:**" as Keyword

   Contents: Les fables de La Fontaine
   People: La Fontaine, Jean de
   Copies: See Catalog

   Contents: Histoire de Babar
   People: Brunhoff, Jean de
   Copies: See Catalog

   Contents: Cin-Blanc
   People: Guillo, Ren
   Copies: See Catalog

   Contents: Lettres de mon moulin
   People: Daudet, Alphonse
   Copies: See Catalog

   Contents: Les fourberies de Scapin
   People: Molière
   Copies: See Catalog

   Contents: La Belle et la Bête
   People: Leprince de Beaumont, Jeanne-Marie
   Copies: See Catalog

7. pôle anonyme La face de maître Pathein, d’après présenté, annoté et commenté par Thierry Revel; traduction de Alain Mij. Paris Larousse 21/08/2013.
   Copies: See Catalog

   Contents: Les allumettes soudées
   People: Sabatier, Robert
   Copies: See Catalog

   Contents: Yvain ou Le chevalier au lion
   People: Chrifien de Troyes. Dacelini, Hfie

---

**Fig. 67.** Extract of the FRBRized BIB-RCAT dataset in Scherzo (Variations/VFRBR)
Below we present the interpretation of a MARC record in FRBR using XC. Figure 68 depicts the initial MARC record. Figures 69, 70 and 71 respectively illustrate the FRBRization into a Work, an Expression and a Manifestation.

Fig. 68. A MARC record displayed in XC web interface
**Fig. 69.** A FRBR Work entity displayed in XC web interface

**Fig. 70.** A FRBR Expression entity displayed in XC web interface
Fig. 71. A FRBR Manifestation entity displayed in XC web interface
7 Acknowledgments

This work has been partially supported by the French Agency ANRT (www.anrt.asso.fr), the company PROGILONE (www.progilone.com/), a PHC Aurora funding (#34047VH) and a CNRS PICS funding (#PICS06945).

References