

## Spot the Difference: Monitoring Data Quality in COBISS.SI

**Branka Badovinac**

IZUM – Institute of Information Science, Maribor, Slovenia

E-mail address: [branka.badovinac@izum.si](mailto:branka.badovinac@izum.si)



Copyright © 2022 by Branka Badovinac This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

---

### Abstract:

*The article outlines a pragmatic dimension-based approach for observing and measuring the quality of data values in bibliographic and authority records. The usefulness of this approach will be demonstrated on the example of the Slovenian Co-operative Online Bibliographic System & Services (COBISS.SI) which has implemented various data quality activities. Two examples of usability of dimension-based approach will be presented: 1) case from the activity identification of record groups with common error patterns and 2) an analysis of the daily bibliographic records production quality monitoring in the period from 2015 to 2020. Although the results show relatively good data quality, it has also been noticed that cataloguers should be more aware of changes in cataloguing practice.*

**Keywords:** data quality, monitoring activities, data quality dimensions, COBISS.SI

---

### 1 Introduction

Data quality is usually embedded in broader evaluation frameworks, in particular information systems, databases, standards etc. In library and information science (LIS) literature, data quality is often included in discussions on 1) end-users' needs, 2) workflow organisation/rationalization, 3) quality from the perspective of cataloguers/metadata specialists, 4) cataloguing technology possibilities, and 5) cataloguing uniformity through standardization.

Data quality measurement activities and bibliographic and authority records evaluation are demanding because of the large number of data elements in data schemas. Data values in records are semantically complex and dynamic. Also, there is no "perfect record" that we can use as a reference to compare with, in order to spot the errors. We rarely have access to the described resource, that is the »item in hand«, and external resources are only partially useful. Besides, some level of subjectivity in decisions on how to describe the bibliographic resources should be taken in account (Hider and Tan, 2008). Moreover, data quality activities

and analyses must also be financially justifiable and feasible, especially since they often require manual methods for error checking and correcting.

Even though a significant amount of research on the topic of data quality in bibliographic systems is already available, what was missing in the activities for monitoring the Slovenian Co-operative Online Bibliographic System & Services (COBISS.SI, <https://www.cobiss.si/en/>) was an integrated approach to data quality. In our case, data value is the value of each data element in bibliographic and authority records according to content and format (schema) standards. What was needed was a framework that would enable defining, measuring and evaluating data quality, but would not be dependent on the changes in standards and formats. A more conceptual level was required that would enable a reasonably simple analysis and understanding of the quality of data entered in the MARC format.

Firstly, this paper will present the basis for the activities of quality monitoring, which are a part of the quality assurance framework in the COBISS.SI system. This will be followed by a presentation of the dimension-based approach, which lists an array of characteristics for describing data quality. The final part of the paper will present two practical examples of using this approach.

## **2 Monitoring data quality in COBISS.SI**

Co-operative Online Bibliographic System & Services or COBISS is designed and maintained by the Institute of Information Science (IZUM) in Slovenia. The COBISS system not only provides support for automation and rationalization of various library processes but is also an organizational model that links individual libraries into a national library information system with a shared cataloguing framework and with shared tools. It includes the COBIB union catalogue/bibliographic database, local bibliographic databases of participating libraries, as well as the database of libraries (COLIB), and the authority records files. COBISS3 is the third generation of software that was developed by IZUM for the COBISS system, and [COBISS+](#) is the name of the online public access catalogue (OPAC) service in COBISS, while mCOBISS is its mobile version. Recently, the dCOBISS repository has also been developed. It is intended for storing the most common types of digital objects and managing digital contents and is fully integrated with all other software applications of the COBISS system.

The two basic principles of COBISS are: shared cataloguing and exchange of records within the COBISS.net network, which includes the national systems from Slovenia, Serbia, North Macedonia, Bosnia and Herzegovina, Montenegro, Bulgaria, Albania, and Kosovo. Every national system uses its own cataloguing code, while the data formats are common. IZUM has developed the following formats: *COMARC/B for bibliographic records* (2022), *COMARC/A for authority data* (2022), and *COMARC/H for holdings data*. They are national implementations of UNIMARC (COBISS Platform, 2022). In Slovenia, Eva Verona's Code and Manual for Compiling Alphabetical Catalogs (1983–1986) is used for cataloguing monographs, and a group of ISBD standards is used for other types of resources. In addition, several amendments have been adopted in recent years to introduce modern cataloguing practices.

The Slovenian system COBISS.SI has been developed since 1984. It includes different types of libraries: national, academic, public, special, and school libraries. Each system member has different groups of end-users, different needs of data management and profiles of

librarians. As can be seen from Table 1, COBISS.SI is a relatively small, yet highly heterogeneous system. In recent years it has been characterised by introducing school libraries to the system and the development of authority files, with subject headings (SGC) being particularly noteworthy. A majority of records in the COBIB.SI database is originally created, by approximately 600 active cataloguers. A total of 301 institutions out of 923 members participate in the shared cataloguing system (COBISS3 Software, 2022). Data from the COBIB.SI shared (union) database and the CONOR.SI authority files for personal and corporate body names are also used by other services within the COBISS.SI system, especially by the Slovenian Current Research Information System ([SICRIS](#)), and the system for remuneration for the public lending of copyrighted works by Slovenian authors.

*Table 1: Selection of indicators of COBISS.SI in 2021 (source: Statistical indicators, 2022; SGC, 2022)*

<i>Number of full members (libraries)</i>	923 (national and academic libraries – 9%, public libraries – 7%, special libraries – 15%, school libraries – 70%)
<i>Number of bibliographic records in COBIB.SI shared database / local databases</i>	5,793,460 / 18,869,145
<i>Increase of bibliographic records in 2021</i>	180,412
<i>Percentage of records by material type</i>	monographs – 56%, component parts – 38%, serials – 3%, performed works, integrating resources, etc. – 4%
<i>Number of authority records</i>	CONOR.SI personal and corporate body names– 1.436.445 SGC subject headings – 63.227
<i>Percentage of bibliographic records linked to CONOR.SI</i>	76% for personal names 4% for corporate names
<i>No. of cataloguers</i>	614 (on 16 June 2022)

The COBISS.SI quality assurance framework uses several record quality control mechanisms, combining standardization, software application and user training. Quality, uniformity and consistency of the local databases and COBIB are ensured by the use of authority control, duplicates control, COMARC format software controls, record editing, global code lists for all standardised data (e.g. countries, languages, UDC), local code lists for all data that is uniform within the library, automatic counters, unique identification control of serials, etc., and above all, by providing systematic training sessions for record creators, who must obtain the official cataloguing permit for their work (COBISS Platform, 2022).

As a part of IZUM's organisation structure of the Bibliographic Control Department, one work group is in charge of some activities related to quality assurance. These activities can be divided into 1) activities before and during data entry, and 2) activities after data entry (monitoring of output data). Within the first group of activities, the cooperation during cataloguer training and software improvement is particularly noteworthy. For an e.g. COBISS3 cataloguing software has a number of automatic validators, which are designed to find basic record errors during the data input phase. The second group of activities consists mainly of:

- regular reviews of records that were created by beginner cataloguers (one year after the after obtaining particular credential for work in cataloguing software),
- yearly reviews of 100 randomly selected records (in co-operation with the national library; review with item in hand),
- cataloguing helpdesk and coordination of error corrections reported by end-users,

- identification of record groups with common error patterns, and
- continuous monitoring of recently created bibliographic records (and linked authority records).

For the analysis of data quality in COBIB.SI, the latter two are of particular interest. Identification of record groups with common error patterns is performed on an occasional basis. When we find errors that are repeated, we perform an analysis on a larger sample of records. The selection of records is usually limited to a certain time period (e.g. the last year or more), and the review of records is carried out without a resource.

The second one, which is also known as daily production quality monitoring, was developed in 2015. Its primary goal is to resolve the errors in recently created records. The monitoring method is based on a sampling of 10% of records that were created on a random day each week, which means that 40 samples of approximately 50 bibliographic records per sample are evaluated each year. The sampled records are then reviewed manually, without a resource in hand, except for online open-access resources. Data is relatively stable at the time of reviewing, as the process is finished within 6 days after the records had been created and sampled.

Cataloguers are notified by email only when a record has at least one major error. A major error can be incorrect data, inconsistency in content and structure, missing or partly missing mandatory data, or redundant data in any data element. At the end of the year, it is checked whether and how errors have been corrected. This is done on a small sample of records.

This kind of monitoring has some general limitations. Only a small number of records is reviewed (approx. 2,500 records per year). Samples do not include records from the National and University Library, which contributes approximately 10% of records to COBIB.SI on a yearly basis. Even though the activity is based on the expert-review method, it still provides a subsample of records that have been checked and re-checked after corrections have been made by cataloguers.

### **3 Data quality dimensions for bibliographic and authority records**

According to a natural or inherent definition, the term “quality” signifies the highest possible degree of capturing a description of the real world, while a pragmatic definition (also known as “fitness to use”) is more common in practice. It is defined as the degree to which a set of inherent characteristics fulfils the consensual requirements. In our case, these characteristics are called dimensions<sup>1</sup>. Dimension signifies a description of data that can be measured based on a pre-existing set of professional goals or cataloguing standards. According to Batini and Scannapieca’s (2016) definition, data quality depends on the number of included dimensions. Dimensions are not independent. They can relate to each other in different ways, such as in the case of an inverse proportion between consistency and completeness.

Based on the typology of errors, we have set up eleven dimensions (Table 2) (Badovinac, 2021), which make up our data quality definition, which is:

Quality data is present in a unique bibliographic or authority record and it is not redundant. It is structurally consistent, structurally complete, semantically accurate, syntactically accurate, value complete, semantically coherent, representationally consistent, current, and may have added value.

---

<sup>1</sup> It can also be called a criterion, requirement, indicator, parameter, etc.

Table 2: Set of dimensions for data quality description in bibliographic and authority records

<b>Dimension</b>	<b>Abbreviation</b>	<b>Description</b>	<b>Precondition</b>
<b>Uniqueness</b>	UNIQ	A record is unique when there is no duplicate at the database level as is required by the reference (standards, etc.).	
<b>Redundancy</b>	REDUN	Data is redundant if it is not required by the reference (standards, etc.).	Uniqueness
<b>Structural consistency</b>	STRUCON	Data is structurally consistent when it is present in the correct data element (field / subfield), and is in the correct order as required by the reference (standards, etc.). On the record level, structural consistency refers to an accurate choice of template or original record.	Uniqueness Redundancy Semantic accuracy Syntactic accuracy
<b>Structural completeness</b>	STRUCOM	Structural completeness requires mandatory presence of data as is required by reference (standards, etc.).	Uniqueness
<b>Semantic accuracy</b>	SEMACC	Data is semantically accurate if it corresponds to a real-world value in the described resource itself or in the reference (standards, etc.).	Uniqueness Redundancy Structural consistency
<b>Syntactic accuracy</b>	SYNACC	Data is syntactically accurate when it corresponds to a string (sequence of characters) as found in the resource itself or in the reference (standards, etc.).	Uniqueness Redundancy Structural consistency Semantic accuracy
<b>Value completeness</b>	VACOM	Data is value complete when all parts of the data from the resource itself or foreseen by the reference (standards, etc.) are present.	Uniqueness Redundancy Structural consistency Semantic accuracy Syntactic accuracy
<b>Semantic coherency</b>	SEMCOH	Data is semantically coherent if it semantically matches other data in the record as is required by the reference (standards, etc.).	Uniqueness Redundancy Syntactic accuracy Structural consistency
<b>Representation consistency</b>	REPCON	Data is consistent if it is represented in the form that is required by the reference (standards, etc.).	Uniqueness Redundancy Structural consistency Semantic accuracy Syntactic accuracy Value completeness Semantic coherency
<b>Currency</b>	CURR	Data is current when it is the lattermost or updated with regard to the described resource itself or the reference (standards, etc.).	Uniqueness Redundancy
<b>Added value</b>	AVAL	Added value data is optional data that contributes to the general record value as is defined by the reference (standards, etc.).	Uniqueness

Firstly, a set of dimensions can serve as a general framework to define data quality in individual bibliographic and authority records. Dimensions can also be turned into control

questions for reviewing the records. Furthermore, the framework can also be used to specify which dimensions are relevant for single data element. This means that the dimensions make it possible to analyse individual error types and to calculate potential errors for individual data elements, which in this case is either data value in the COMARC subfields or in the indicators. The sum of all dimensions that refers to certain data element can be also the sum of all potential errors of data value of that data element. In the case of COMARC/B in the software COBISS3, quality of data value in subfield 100c (Publication date 1) can be described by two dimensions: semantically accurate (SEMACC), and consistent with other data elements (SEMCOH). This in turn means that there are two possible error types in subfield 100c. Following this principle, it is possible to speculate that the number of potential errors in bibliographic and authority records is very high.

Measuring should be rational and include only those data elements that are important in a given context. As dimensions on their own do not provide any methods for quantitative measurement, one or more metrics have to be associated with them as separate properties. For each metric, several measurement and evaluation methods can be applied (see e.g. Loshin, 2011).

Measuring and evaluation of data quality within the activity of *identification of record groups with common error patterns* is based on data sampling and assessment possibilities, whereas for the activity of *continuous monitoring of the recently created bibliographic records* a dimension-based analysis can only give a general insight into the quality of bibliographic records. To calculate the average number of errors in individual records, it is also necessary to know the number of all possible data elements in an individual record as well as the number of all potential errors in an individual data element pertaining to a particular record that has been reviewed. The weight or the importance of individual dimensions should also be taken into consideration.

## **4 Usability of dimension-based approach: two examples**

### **4.1 Identification of record groups with common error patterns: syntactic accuracy in title proper and subject terms**

The importance of typographical errors on the searchability of resources has been highlighted in many studies and discussions (see e.g. Beal and Kafadar, 2004). In 2016, the use of the Slovene language spell-checker was enabled in the COBISS3 software. As a consequence, we wanted to check the rate of typos before and after the introduction of the spell-checker. We chose two data elements: *title proper* (in COMARC/B, subfield 200a) and *subject terms* (in COMARC/B, subfields 600–610, and 960–969). These are also essential access points according to the *Statement of International Cataloguing Principles* (Galeffi, A. et al., 2017).

As opposed to errors that are related to the semantic accuracy (SEMACC) dimension, i.e. the dimension where the actually data value is not recognised, bad data within the syntactic accuracy (SYNACC) dimensions is recognised or is partially understandable. Such errors are e.g. superfluous spaces, punctuation, swapped letters.

Two checks of data in the Slovene language were performed. The sample included records created between 2014 and 2018. For title proper, 370,192 data values (fields 200a) from 370,192 records were sampled. For subject terms records were sampled with a string of

characters<sup>2</sup> in fields 600–610, which were selected following an analysis of frequent subject headings and frequently mistyped words in Slovenian. The sample included 296,049 subject headings from 231.666 bibliographic records. Each data value was checked manually, and the cataloguer was informed about any detected errors.

The analysis showed that for 666,241 data values from 601,858 bibliographic records there were only 1,923 cases of mistyped data, 1,150 errors (in 1072 records) were in the title proper, and 773 errors (in 763 records) were in subject headings. Also, we established that the spell-checker has a positive impact on the data in the title, whereas in the case of subject headings, the effect is a bit smaller (Figure 1). For the latter we assumed that the cataloguers did not activate the spell-checker when entering the data.

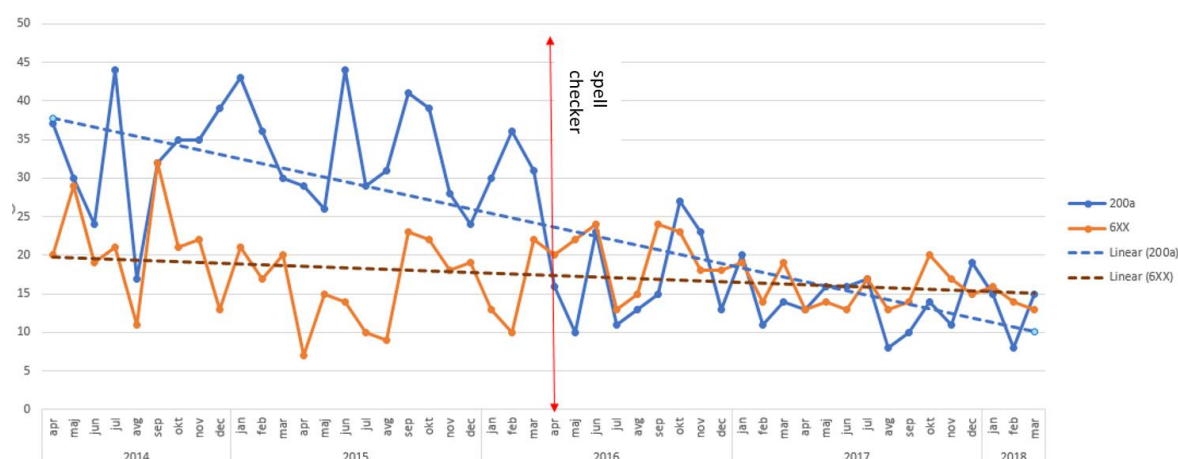


Figure 1: Typographical errors in the title proper and subject terms by time range (2014–2018,  $n = 1.923$ )

#### 4.2 Continuous monitoring of recently created bibliographic records and authority records: analysis of results in 2015–2020

From June 2015 to the end of 2020, the daily production quality monitoring method was used to capture 133,027 bibliographic records, of which 13,478 were sampled (Table 3). The manual review included bibliographic records and its linked authority records. The sampling proved to be reasonably adequate, as we captured the general characteristics of yearly indicators of shared database COBIB.SI (Dornik, 2021).

Record assessment showed that about 44% of the bibliographic records with the corresponding authority records were error-free. Of the remaining, 32% had at least one minor issue, while 24% records required corrections due to one or more major error(s). For the latter, we also sent a message to the cataloguers to check the data and correct it if necessary. On that basis, it can be estimated that, according to current cataloguing practice, about 70% records are good or good enough.

Table 3: Summary of daily bibliographic records production quality monitoring 2015–2020

Indicator/Year	2015*	2016	2017	2018	2019	2020	Total
No. of captures	24	50	40	42	41	40	237
No. of captured bib. records	12,445	29,630	20,051	23,261	23,609	24,031	133,027
No. of sampled bib. records	1251	2956	2202	2323	2347	2398	13,478

<sup>2</sup> These were subject terms containing the character string »svet«, »knji«, »love« ali »sti«.

<b>Indicator/Year</b>	<b>2015*</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>	<b>Total</b>
<i>No. of records without errors**</i>	672	1461	920	934	1052	837	5876
<i>No. of records with major error**</i>	388	691	509	518	771	599	3180
<i>No. of records with minor error**</i>	186	747	750	828	477	922	4204
<i>No. of not evaluated records</i>	5	57	23	43	49	40	217
<i>No. of sent messages</i>	326	260	435	452	411	486	2370
<i>No. of data elements with error (bib. records)***</i>	873	2389	2012	2131	1917	2112	11,434

*\* June–December 2015. \*\* Errors in the corresponding authority records are also included. \*\*\* It includes one data element or a combination of several data elements, as well as a field level and a general note.*

Next, every error description pertaining to a particular dimension was coded. There was 11,434 error in 6,241 bibliographic records. Poor data according to the impact to the data quality dimensions in bibliographic records by year is shown in Figure 2.

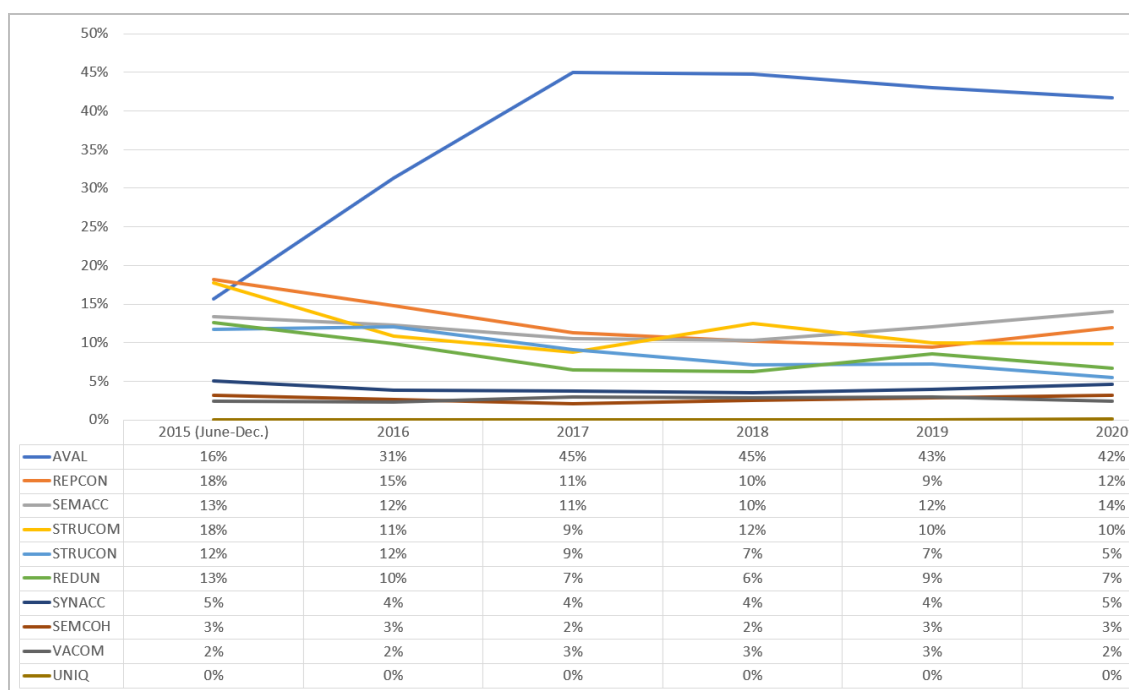
The first finding from the results was that many of the errors coincide with the implementation of new rules in cataloguing. The next characteristic of the observed time period was a very slow decrease in the number of recommendations for entering additional, non-mandatory data. Most of the reviewed notes related to the lack of recommended data (AVAL, 38.9%). From the aspect of catalogue functionality on the shared bibliographic database level, we pointed to mainly missing recommended data related to the implementation of facet navigation in COBISS+. We detected a slightly higher awareness when entering data for data elements for subject headings as only approximately 9% of records had a subject heading missing. We also noticed that a certain part of missing data was related to bibliography management for researchers for the purposes of the CRIS system.

The percentages of missing mandatory (STRUCOM) and incomplete data (VACOM) in the checked records was relatively small (13.6% in total). On the other hand, it is interesting that 8% of redundant data (REDUN) were recorded, which stems from lack of knowledge of cataloguing rules and changes in cataloguing practice. It can be assumed that one of the reasons was also the method of creating new records based on adopting and changing a similar record.

Poor data based on semantic accuracy (SEMACC) and syntactic accuracy (SYNACC) are most problematic from the aspect of data quality. In the observed period, there were 16% of such errors in total out of all the recorded shortcomings. A more precise analysis shows that there were not many cases of incorrect data that would harshly limit the findability of the resource. More problematic were errors affecting the access to online resources (e.g. errors in URL).

Similarly to accuracy, structural consistency is also important. It represented 9% of all poor data. Mostly problematic was the data in the »Title and statement responsibility field« (COMARC/B, field 200) where errors of putting values in wrong data elements is noticed (e.g. additions to the title, responsibility). This kind of errors cannot be prevented by software controls. The cataloguer must therefore be familiar with the rules of using individual data elements and how their order affects the importance of the data. A cataloguer must also pay attention to which structure of data elements to use for the description of the bibliographic resource which is the basis for identifying the resource.





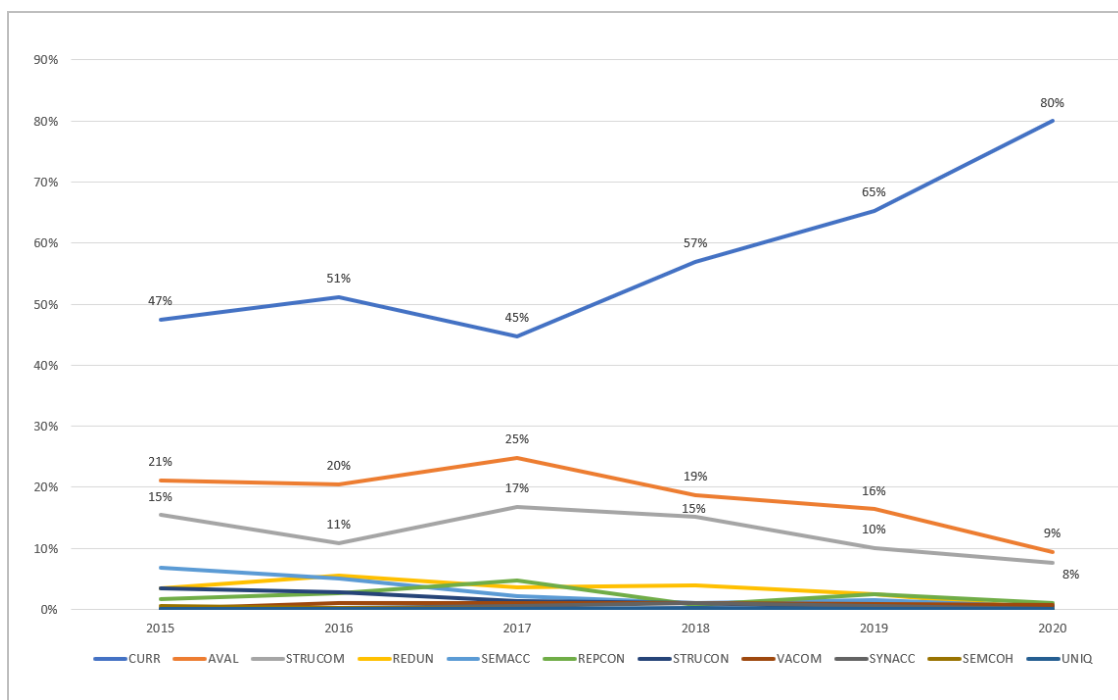
*Legend: UNIQ – Uniqueness, REDUN – Redundancy, STRUCON – Structural consistency, STRUCOM – Structural completeness, SEMACC – Semantic accuracy, SYNACC – Syntactic accuracy, VACOM – Value completeness, SEMCOH – Semantic coherency, REPCON – Representation consistency, CURR-Currency, AVAL – Added value*

*Figure 2: Poor data according to the data quality dimensions in bibliographic records by year (2015–2020, n = 11,434)*

The entry of data to the COMARC format, which is based on the structure of data representation based on the ISBD standard, sometimes demands to entry of same type of data more than once. This can cause the error of data incoherency. As the results of data analysis for the dimension of semantic coherency show (SEMCOH, 2.6%), the errors were mostly related to data elements linked to the publication year.

Representation consistency (REPCON) was, with a percentage of 12%, the second largest group of poor data. However, these errors relating to formatting data do not have a significant effect on the end-user. Lastly, a very small number of duplicate records (UNIQ) was discovered, as the COBISS3 software performs a check when saving the record.

In the linked authority records, 3,033 cases of poor data were detected. It was discovered that, despite the recommendations, cataloguers with the editing credentials, do not update the authority records (Figure 3). Over two thirds of remarks were related to the recommendation for updating the whole authority record (CURR). These are cases for which we estimate that the record contains enough data to be completed. The remaining shortcomings were related to the 908 different authority records where the recommended (AVAL, 16.8%) and other data (STRUCOM, 11.5%) was missing. In relation to the share of poor data, there were only a few other shortcomings.



*Legend: UNIQ – Uniqueness, REDUN – Redundancy, STRUCON – Structural consistency, STRUCOM – Structural completeness, SEMACC – Semantic accuracy, SYNACC – Syntactic accuracy, VACOM – Value completeness, SEMCOH – Semantic coherency, REPCON – Representation consistency, CURR–Currency, AVAL – Added value.*

*Figure 3: Poor data according to the data quality dimensions in authority records by year (2015–2020, n = 3,033)*

## 5 Conclusion

This article has presented the dimension-based approach that was applied in the monitoring activities carried out to check and correct the errors in bibliographic and authority records in COBISS.SI. A dimension-based approach to describing data quality has the potential to provide a more systematic understanding of data quality, as well as a more transparent presentation of results. It enables to regularly and systematically detect and solve individual problems as well as systemic deficiencies.

Even though defining quality is crucial, in order to carry out individual activities, the methods for capturing, measuring and evaluation are also important. For example: monitoring data values' semantic accuracy still remains a methodological and an implementational challenge that we will have to address in the future.

Furthermore, it is necessary to consider which errors and how many of them are (relatively) acceptable, since it is impossible to expect absolute perfection of all data values. By doing so, an insight would be provided into how successful and effective some of the activities are. The approximate framework for acceptable quality standards would be set. Some pointers in this direction are provided by the chronological overview of quality monitoring and measuring. Current results of monitoring activities suggest that an acceptable ratio of good to bad data (or records) could be around 80:20.

The results of the described cases indicate that the recent bibliographic and authority records in system COBISS.SI are relatively good in view of the valid cataloguing practice. Special attention must be paid to activities that are carried out before and during data input. Finding and correcting errors is known to be significantly more expensive than controlling data input

(Redman, Fox in Levitin, 2009). In addition to improving the standards, software and training, cataloguers should also adopt an awareness to keep up with the changes in cataloguing practice.

## References

- Badovinac, B. (2021). Defining Data Quality in Bibliographic and Authority Records: A Case Study of the COBISS. SI System. *Cataloging & Classification Quarterly*, 59 (4), 355–377.
- Batini, C. and Scannapieco, M. (2016). *Data and Information Quality: Dimensions, Principles and Techniques*. Berlin: Springer.
- Beall, J., & Kafadar, K. (2004). The Effectiveness of Copy Cataloging at Eliminating Typographical Errors in Shared Bibliographic Records. *Library Resources & Technical Services*, 48 (2), 92–101.
- COBISS Platform – Co-operative Online Bibliographic System and Services* (2022). Maribor: IZUM. Available at: <https://www.cobiss.net/cobiss-platform.htm> (16. 6. 2022).
- COBISS3 Software* (2022). Maribor: IZUM. Available at: <https://www.cobiss.si/en/c3/> (16. 6. 2022).
- COMARC/A*, 2022. Maribor: IZUM. Available at: [http://home.izum.si/izum/e\\_manuals\\_html/COMARC\\_B/slv/ch1.html](http://home.izum.si/izum/e_manuals_html/COMARC_B/slv/ch1.html) (16. 6. 2022).
- COMARC/B*, 2022. Maribor: IZUM. Available at: [http://home.izum.si/izum/e\\_manuals\\_html/COMARC\\_A/slv/ch1.html](http://home.izum.si/izum/e_manuals_html/COMARC_A/slv/ch1.html) (16. 6. 2022).
- Dornik, E. (2021). COBIB.SI: prirast bibliografskih zapisov v letu 2020 [COBIB.SI: increment of bibliographic records in 2020]. *Organizacija znanja*, 26 (1–2), 2126007, Available at: [https://www.cobiss.si/OZ/PDF/OZ\\_2021\\_1\\_2\\_final/2126007\\_Dornik.pdf](https://www.cobiss.si/OZ/PDF/OZ_2021_1_2_final/2126007_Dornik.pdf).
- Hider, P. in Tan, K. (2008). Constructing Record Quality Measures Based on Catalog Use. *Cataloging & Classification Quarterly*, 46 (4), 338–360.
- Galeffi, A. et al. (2017). *Statement of International Cataloguing Principles* (2017). Den Haag: IFLA. Available at: [https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp\\_2016-en.pdf](https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-en.pdf) (16. 6. 2022).
- Loshin, D. (2011). *The Practitioner's Guide to Data Quality Improvement*. Amsterdam: Elsevier.
- Redman, T. C., Fox, C. in Levitin, A. (2009). Data and Data Quality. In *Encyclopedia of Library and Information Sciences*. New York: Taylor and Francis (pp. 1420–1431).
- SGC – COBISS.SI General List of Subject Headings* (2022). Maribor: IZUM. Available at: <https://www.cobiss.si/en/sgc/> (16. 6. 2022).
- Statistical indicators* (2022). Maribor: IZUM. Available at: <https://www.cobiss.si/en/indicators/> (16. 6. 2022).