

Digital Research Collection – Principles and the Case of the Ethnological Collection of Research Reports

ANDREJ GOGORA



DOI: <https://doi.org/10.31577/SN.2023.2.13> © Ústav etnológie a sociálnej antropológie SAV, v. v. i. © 2023, Andrej Gogora. This is an open access article licensed under the Creative Commons

Andrej Gogora, Institute of Ethnology and Social Anthropology of Slovak Academy of Sciences, Klemensova 19, 813 64, Bratislava; e-mail: andrej.gogora@savba.sk

The article is localized at the intersection of digital curation, archival studies, digital humanities, and the documentation practice in ethnology. The primary aims are 1. to analyze the five selected principles of conceptual preparation and practical building of digital research collections; 2. to exemplify these principles on the concept and strategy of digital conversion and computer processing of the ethnological Collection of research reports (CRR) at the Institute of Ethnology and Social Anthropology SAS; 3. to evaluate and to comment on the selected principles of research collections based on CRR exemplification. The first aim includes defining essential terms, principles analysis (principle of development policy, description, digital curation, contextual mass, and scholarly contribution) and highlighting the specificity of thematic research collections. The second aim involves in-depth exemplifying conceptual, methodological, curatorial, and practical processes through which CRR becomes a digital research collection. The third aim includes a retrospective evaluation of problematic aspects of selected principles from the point of view of the CRR example. The article's contribution will be to improve the knowledge of professionals in ethnology about the theoretical foundations and curatorial management of digital research collections.

Keywords: thematic research collection, ethnographic digital resources, ethnographic research reports, digital curation management, digital curation and research environment life cycle

How to cite: Gogora, A. (2023). Digital Research Collection – Principles and the Case of the Ethnological Collection of Research Reports. *Slovenský národopis*, 71(2), 107–125. DOI: <https://doi.org/10.31577/SN.2023.2.13>

Introduction

The intention to write the article emerged from experience during the author's work at the Research collections department at the Institute of ethnology and social anthropology Slovak Academy of Sciences (IESA SAS)¹. After he joined the workplace in 2017, he took over the department under challenging conditions. Despite several important outputs realized in the past², there had been a long-term neglect of the latest trends in digital curation and archival practice, unfamiliarity with the benefits digital collection offers for research practice, and a state of emergency of the physical documents.

The article's content, regarding discipline, can be defined at the intersection of digital curation, archival studies, digital humanities, and the documentation practice in ethnology. It follows that it represents a mixture of theoretical, methodological, practical, organizational, and partly institutional knowledge and approaches. It is a balanced mixture because the extensive use of practical and methodological knowledge supports the theoretical findings about digital collections. The primary aims of the article are:

1. to analyze the selected principles of conceptual preparation and practical building of digital research collections;
2. to exemplify these principles on the concept and strategy of computer processing of the ethnological Collection of research reports (CRR) in IESA SAS;
3. to evaluate and to comment on the selected principles of research collections based on CRR exemplification.

The first aim includes, besides principles analysis, defining basic terms; highlighting the specificity of thematic research collections (compared to collections as such). The second aim consists of a CRR exemplification, an explanation of the theoretical, methodological, and practical consequences of the particular case that meets the requirements of the digital research collection. The third aim includes a retrospective evaluation of problematic aspects of selected principles from the point of view of the CRR example.

In addition, it is crucial to point out that the article focuses on the issue of research collections as such (digital or non-digital), primarily not on the entities of which collections are composed. Digital objects, metadata information, database, and other

1 The research collections department has been operating in the IESA SAS since 1953. It includes extensive textual, pictorial, and multimedia documents and provides documentation, information, and technological support for the scientific and research activities of the staff: <https://uesa.sav.sk/en/research-collections/>.

2 First of all, it is an extensive online image database (more than 118.000 records) developed within an international project *Preservation and Enhancement of Folk Culture Heritage in Central Europe - ETNOFOLK* (2011–2014): <https://uetetnofolk.eu/index.php/main/about>.

entities will be investigated, but mainly to support generating a comprehensive view of digital research collections. Hence, the contribution will be to improve the knowledge of professionals in the ethnology of the theoretical foundations and management of digital research collections. The realization of this benefit is essential for the following reasons:

1. in the last two decades, the incidence and usability of digital research collections in the humanities have constantly been growing (Palmer, 2004: 348–365);
2. due to the persistent unsatisfactory condition of Slovak digital research collections in ethnology, the need for a sustainable building is urgent (Zajonc, 2006: 30–47; Panczová, Kiliánová, Kubisa, 2021: 71–73);
3. the use of digital research collections can facilitate research practice (Hughes, Constantopoulos, Dallas, 2015: 161–162).

1. Terminology

Some humanities scholars may need help understanding the terms from computer science or library and information science, so the next part deals with their definitions. The term digital collection belongs to the broader concept of **digital resources**, which is any information material created or converted into electronic format (e.g., digital audio, images, text, video, multimedia, database, bibliography, editions).

A **digital collection** is a group of digital objects selected according to one or several attributes, arranged and maintained to ensure their access, browsing, preserving, and querying for the end-user. The digital collection can be a true copy of a physical one. However, it can effectively mix and remix items from different collections, subcollections, or other types of digital resources (Sabharwal, 2015). The digital and analog collections differ in format and content management.

The feature that enables this creative combinatorial curation process is that it operates with **discrete items** that are not exclusive parts of a larger body of material (NISO, 2007: 2–4; Carignan et al., 2007: 1). Discrete items can be easily isolated from the original context and accumulated to form a new series. The digital collection can then virtually aggregate various objects from various sources, non-digital or born-digital.

The term that directly relates to the interrelation of discrete items is **digital research collection**, a unique collection of primary sources and other materials gathered by scholars to support research on a theme (Fenlon, 2017). Unsworth extensively defined it as a large-scale interdisciplinary electronic resource, structured but open-ended, diverse in datatypes but thematically coherent, and designed to support research purposes (Unsworth, 2000). Miller (2019) points out that this term is often incorrectly confused with digital collection. Fenlon underlined its particularity: “...if thematic research collections are a special type of collection, they may place an

even higher demand on curatorial intent: not only that it exists, but that it be of a particular sort – an intention *to support research on a theme*” (Fenlon, 2017: 22). Regarding typology, research collections are diverse in terms of subject matter, scope, technological requirements, functionality, authorship, and method of presentation. Some collections ambitiously focus on entire historical periods, a nation’s interests, or broad areas of knowledge, and others accumulate the work of a single author or time-lapse notes. Some offer a user interface with advanced statistical data processing tools, others only basic search functions. Most well-known research collections are institutional multi-author large-scale projects, but personal thematic collections are frequent (Fenlon, Senseney, Green, Bhattacharyya, Willis, Downie, 2014).

2. Principles of digital research collection

To achieve the first aim, selecting a set of principles was necessary. Searching for resources proved that composing it from subsets that represent both practical and conceptual points of building digital research collections is appropriate. The first three selected principles relate to digital collections as such, and they are more technological (NISO, 2007). These are basic principles, but it is reasonable to include them regarding the discussed example. The other set of two principles comprehensively reflects the essential aspects of digital research collections (Palmer, 2004; Unsworth, 2000). There are other lists, but their principles are not helpful or already present in the selected set (Pennock, 2013; NSLA, 2009).

IESA SAS staff has been building CRR since 1946, and it contains over 1.500 documents (a total of 158.000 items in various formats) recorded in the ethnological field and archival research. Thematically, it comprehensively covers traditional cultural heritage throughout Slovakia (Table 2, Table 3), however, from the beginning it was mainly an effort to capture the disappearing phenomena of traditional folk culture (Prandová, 1982). From a methodological point of view, they were carried out as thematic and regional field research (either as the researcher’s direct observation, or the aggregation of various secondary written and visual materials). There are several reasons to use it as an example of research collection. Firstly, it was systematically built and documented over a long period by experts in ethnology. Secondly, the collection is currently in an advanced state of analog-to-digital conversion, which will help illustrate the principles in their application and at different levels of development. Efficient strategies will be confirmed, mistakes will be pointed out, and further methodical instructions will be more clearly defined.

Format	A4 paper size	A5 paper size	other formats	photograph	drawing
Number	83 698	70 540	2 062	1004	719

Table 1. Types and numbers of formats of CRR’s documents

2.1. Development policy

The principle of development policy articulates the conditions for establishing, acquiring, maintaining, and provisioning a collection: “A digital collection should be the result of careful selection and collection policies. In addition to applicable copyright and technical considerations, the scope and nature of the materials, as well as the intended audience, should be taken into account” (Carignan et al., 2007: 11). The policy further includes digitizing guidelines, material removal criteria, and rules for creating collections. Notably, the institutional mission should primarily determine the setting of these variables (both technological and conceptual). This mission-driven nature of development policy ensures that it is distinct from technicalities such as a digitization manual, as is often the practice (Cole, Jackson, Palmer, Shreeves, Twidale, Zavalina, 2006: 4). The creation and use of the collection should consider and support the organization’s primary values and goals. The right development policy will ensure that the collection provides users with professionally prepared and curated resources that meet institutional needs (Duranti, Franks, *Eds.*, 2015: 142–145). Concerning a research institution, it is clear that its collections should support the research mission and scientific or educational conception.

In the case of analog CRR, a particular document defines its development policy (Prandová, 1982). It contains the essential elements of a development policy, including an explicit institution mission statement and its implementation using documentation, preservation, and scientific use of field research reports. However, it is called an “orientation handbook”; it does not contain complete documentation of instructions for creating collection catalogs; and was published 36 years after the establishment of the collection. In addition, much knowledge and practices were transmitted by mouth. The following two main issues have emerged in the long-term practice:

1. documents that did not meet the formal and content criteria of research reports were often added to the CRR;
2. CRR was initially called the *Text Archive*, so the formal criterion for document acquiring was textuality.

These issues have caused the analog CRR to have high qualitative and content inconsistency. Despite the declared textuality, many documents contain mixed formats such as photos, drawings, textile fabrics, and music sheets. Although implementing the policy was problematic and the possibilities to curate physical collections were limited, massive documentation and preservation of domestic ethnologically valuable material were realized.

Considering these issues, the conversion of this uncurated mass into an up-to-date digital research collection turned out to be very problematic in 2017. The process began after a departmental generational change. Hence, an initial audit and manual transcription of the accession book had to be done first (including complete chemical treatment in an ethylene oxide chamber and archival repackaging). Only then did the reworking of the development policy towards the digital curation policy document

begin. In short, the new policy document preserved the original entry documentation procedures; added digitization, post-processing and OCR standards; defined document selection criteria; outlined potential user groups; includes an anonymization strategy for GDPR regulation; and enriched the mission statement with digital curation and research environment life cycle (research project cooperations). In addition, the formal criterion of textuality was eliminated; some collections were extracted and renamed (e.g., Collection of F. Wollman, sub-collection of questionnaires), and documents that did not meet the criteria of research reports were singled out. Detailed time phasing was worked out, and the documentation of the entire process was continuously recorded.

In conclusion, the first principle of development policy requires clear objectives and detailed knowledge of the items to avoid unsolicited additional changes in the policy that can disrupt the management and quality of the collection. However, in the case of collections created by digital conversion of long-term, massive, neglected, and heterogeneous funds, such as CRR, it is strategic to consider the degree of commitment to development policy in advance. During the processing of the CRR, several complications arose that were difficult to foresee, e.g. several lost documents, duplicate numbering of documents, key data in the catalogs are written by hand, controversial discussion about ethnological subject systems. Additional changes in the document are possible due to the current needs and capacities or to improve the fulfillment of the mission statement.

2.2. Collection description

The second principle says that description provides information about the collection and its items. Such a description is called metadata, concisely defined as data about data. In defining this term, some scholars emphasize the epistemic nature of metadata when claiming it is a statement about a potentially informative object (Pomerantz, 2015). Others stress the systematic composition of its units when defining it as a structured description of the attributes of an object (Gill, 2008). Others focus on the practical application when a user through metadata is browsing, querying, managing, and preserving the collection. So its purpose is to render the semantic map that provides the user with easy navigation when exploring the structured digital object.

There are five basic types of metadata: **descriptive** (provides content description of an object such as author, title, locality, subject), **administrative** (helps to manage an object such as acquisition information, creator, rights), **structural** (provides information about the resource such as chapters, number of pages, headings), **preservation** (relates to the information about long-term retaining of an object such as digital file information, scanning resolution), **use** (statistical information regarding the use of the object such as user tracking, search logs, number of downloads). Finally, it is worth reminding, due to the third principle, that metadata are discrete items or objects. They are not functionally fixed and can be processed regardless of the source data.

The physical description source in CRR is the accession book (the mixture of descriptive, administrative, and structural metadata) and subject and locality card catalogs (primarily descriptive ones). During the creation of the development policy, it was problematic to plan the automated digital conversion of the metadata source. The reason was that the accession book and the crucial data in the subject card catalog were written entirely by hand. The department owns a feed scanner for the massive digitization of such cards (approximately 14.000). However, despite the progress of handwritten text recognition (HTR) technology in recent years, the process aimed at automatic text data extraction would be technologically and time ineffective – the handwriting was frequently illegible, the paper was often damaged and written on both sides, the card’s segmentation was inconsistent for automated document layout analysis. Both sources were transcribed (the accession book contains localities data) into the spreadsheets containing data on document ID, title, author, number and format of items, language, period of research, localities, subjects, year, and method of acquisition (Table 4).

	District	Incidence
1	Brezno	339
2	Krupina	275
3	Rožňava	188
4	Veľký Krtíš	170
5	Michalovce	155
6	Levice	129
7	Bardejov	126
8	Trebišov	116
9	Liptovský Mikuláš	110
10	Čadca	97
11	Rimavská Sobota	96
12	Trenčín	90
13	Malacky	87
14	Bánovce nad Bebravou	82
15	Prešov	80

Table 2. The fifteen most frequent Slovak districts in which field research for CRR were carried out according to the locality catalog (the total number of districts is 84, individual municipalities are located within districts, while CRR contains a total of 5359 municipalities’ incidence).

When considering well-prepared metadata, there are three conditions in the foreground: completeness, standardization, and good-structuring. The digital conversion of CRR metadata is currently in the final stages of implementation but can be evaluated according to these conditions. The **completeness** (description should not be ambiguous or fragmentary, it should comprehensively describe a resource) of CRR metadata is rather a matter of consistently applying the archival principle of provenance (Lorenzini, Rospocher, Tonelli, 2021). Due to the decision

to respect the original metadata, the responsibility for completeness is transferred back to the creators of the analog CRR catalogs. The reason for the metadata adoption was that they were systematically and professionally created over a long period by the department staff in coordination with the authors of documents. With such a mass of documents, the effort to generate new metadata records would be unmanageable regarding time capacity and expertise. The original administrative and structural metadata were sufficient, and descriptive ones need a revision of subject taxonomy. These issues fall under the following condition.

Regarding the **standardization** (description should not be arbitrary, it should support a standard recording and understanding of data), due to the acceptance of original metadata, there is a problem with ambiguity, neglected management and outdated hierarchical system of subject classification (Žatko, 1953). This phenomenon fundamentally disrupts the requirement of shared understanding and of building a controlled lexicon, thus the condition of standardization. For example, the subject classification system was frequently and uncontrollably changed; due to the terminological dynamics, many entries in the subject catalog are vernacular and mutually synonymous; one item of material culture is often named in several dialects in the subject catalog; the documents often contain judgmental (misogynic), racist (mostly anti-Semitic), ideologically and politically unacceptable expressions (totalitarian fascist and communist rhetoric); many original places from the location catalog do not exist or have been renamed several times. This condition is also weakened by the original rigid hierarchical structure of subjects, which was suitable for describing the cultural phenomena of that time. However, it fails when it is necessary to include and categorize a new phenomenon (Popelková, Zajonc, 2000) or to meaningfully reconstruct it into a linear subject metadata system. The strategy of preserving the original subject catalog remains valid, but with the requirement of terminological refinement and modification of the taxonomy according to the practical, technological and semantic needs of the digital research collection management and user experience. There is still an active discussion about these problems, which requires the joint involvement of experts from ethnology and digital curation.

	Category	Incidence
I.	Employment of the population, production, transport, trade	3932
II.	Food (in general)	208
III.	Clothing	1582
IV.	Construction (settlement, types of housing, furniture)	910
V.	Hygiene and cosmetics	79
VI.	Social organizations, groups, social and family relationships, customs	590
VII.	Customs, social practices, rituals, festive events	1334
VIII.	Knowledge of the people, religion	396
IX.	Folk fine arts	111

X.	Folk theatre, narrative folklore	334
XI.	Singing, music, dance	280
XII.	Auxiliary disciplines	239
XIII.	History of ethnography and folk culture	120

Table 3. Incidence of subject headings within the main categories from the original CRR subject catalog (the main categories contain three hierarchically subordinate levels and the corresponding coding)

Regarding the **good-structuring** (description should not be randomly organized, it should ensure a consistent scheme in favor of data sharing and interoperability) (Cole et al., 2006), the analog CRR metadata entries had a stable shared structure (though archive schema type not specified) that sufficiently met the needs of work with the collection at that time. For the current needs, the necessity of reconstructing the established structure to the standardized Dublin Core Metadata Initiative (DCMI) terms emphasizes the CRR development policy. This modification was implemented for all types of metadata, individual terms, and corresponding value formats in CRR (Table 4). This strategy will ensure the interoperability of the CRR, meaning the potential linkage and efficient exchange of metadata with other collections within the department (textual or visual) or with external collections and systems, thus potentially wider use for research purposes.

Dublin Core term	Definition
Term Name: title	A name given to the document
Term Name: alternative	An alternative name for the document
Term Name: identifier	An unambiguous inventory number (CRR0001, CRR0002, ...)
Term Name: creator	An entity/author primarily responsible for making the document
Term Name: publisher	An institution responsible for making the document available
Term Name: tableOfContents	A list of subunits of the document
Term Name: type	The nature or genre of the document
Term Name: subject	A topic of the document (subject catalog controlled vocabulary)
Term Name: location	A spatial region or named place
Term Name: language	A language of the document (ISO 639-2)
Term Name: date	A point or period of time of creation of the document
Term Name: dateSubmitted	Date of submission into a collection (ISO 8601)
Term Name: accrualMethod	The method by which the document is added to a collection
Term Name: format	The file format, physical medium, or dimensions of the document

Term Name: accessRights	Information about who accesses the document and security status
Term Name: bibliographic Citation	A bibliographic reference for the document

Table 4. The metadata structure of the main CRR terms adapted according to DCMI metadata terms: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>.

The fulfillment of these three main conditions will bring the following benefits to CRR administrators and users: it will make searching and filtering using metadata available to ordinary users (previously possible through a face-to-face request to the collection administrator); it will enable more precise and more accurate searching using metadata, and provide a more comprehensive semantic map of the collection; will allow content filtering using a combination of multiple requests; will enable productive research data sharing; and finally it will enable advanced digital curation of the collection.

In conclusion, the description principle is crucial for implementing digital curation and the subsequent research use of the collection. It is necessary to respect current metadata creation trends and adopt structured schema and standardized content. However, many cases, such as CRR, are forced to take over the original metadata source in the digital conversion process and implement very particular descriptive content and format. For example, the narrow thematic specification of a document’s description schema may problematize the requirement of interoperability of metadata systems, and the use of outdated or colloquial folk expressions may decrease the sharability of controlled lexicons. Applying the principle of description largely depends on the type and thematic focus of the research collection.

2.3. Digital curation

The term *curation* originally referred to guardianship and healing then applied to the function of a library custodian, and only then to the care of digital objects. Digital curation synthesizes the first two principles by emphasizing the entire data processing life-cycle. A leading institution in the field³ defines it as managing and preserving digital research data over the long term (Abbott, 2008). The extensive definition reads: “Digital curation is concerned with actively managing data for as long as it continues to be of scholarly, scientific, research, administrative, and/or personal interest, with the aims of supporting reproducibility, reuse of, and adding value to that data, managing it from its point of creation until it is determined not to be useful, and ensuring its accessibility, preservation, authenticity, and integrity over time” (Harvey, Oliver, 2016: 8–9). Then, it is clear that digital curation is a broader term than digital archiving or preservation, which it encompasses but is often mistakenly confused with.

3 The Digital Curation Centre (DCC): <https://www.dcc.ac.uk>

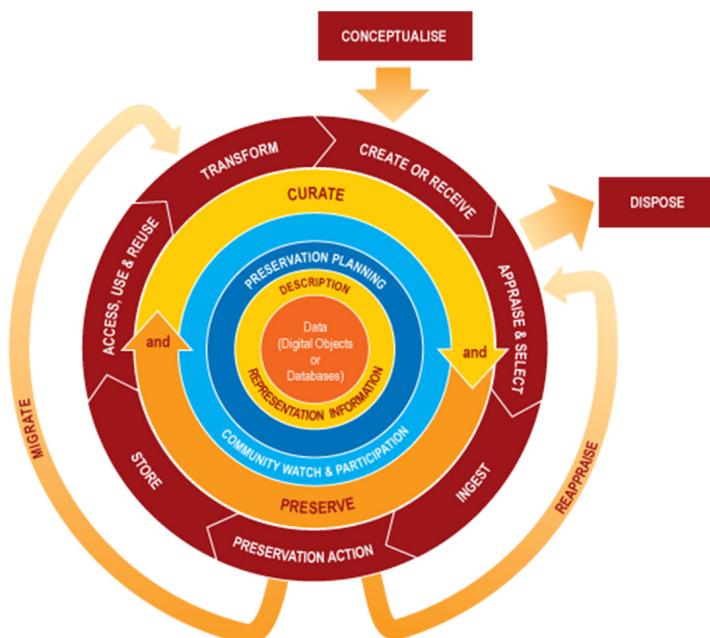


Figure I. DCC curation life-cycle model.
Credit: Digital Curation Centre, University of Edinburgh.

Three key features can specify digital curation. **Temporality** (Figure I. clearly shows the life-cycle approach, which emphasizes the detailed planning of digital curation work procedures /Higgins, 2008/ and the recirculated lifetime of digital objects from their creation to their removal – the collection item is not a static terminated digital object). **Inclusiveness** (digital curation includes a wide range of unrestricted practical and conceptual actions /NISO, 2007: 9/ such as creating, enhancing, correcting, adding, linking, enriching, describing, and transferring with any digital object). **Sustainability** (combinatory and dynamism of curatorial data management /Giaretta, 2011: 38/ ensures that digital objects can be continually used and recycled now or in the future, which increases their value and sustainability).

Firstly, what has been said about digital curation of collections in principle applies to thematic research collections, including CRR. As mentioned above, the CRR was originally called *Text Archive* and was recently renamed. It was not just a formal act but related to a change in approach and work with documents, corresponding to digital curation rather than archiving. The contrast between archival and digital curation principles is illusive here, in this case it is rather a problem of long-term neglect of appropriate archival practice (taking care of the physical condition of documents, mapping provenance, cooperating with creators, working with designated communities, and customizing selection, preservation, and access principles) and of interest in the application of new trends in collection management. The inclination towards inclusive processes of curation in 2017 is also related to the fact that there

was no intention to create a mirror version (it means ordinary digitization) of the *Text Archive*. After all, curating such a massive collection, diverse in format and content, proved necessary.

Within the computer processing of the CRR, it is possible to identify the implementation of curatorial strategy: the development policy conceptualizes the life-cycle of digital objects (explicitly formulated *Principles of document preparation* for accepting new acquisitions); unsuitable and unusable items were discarded or set aside (128 documents); all digitized documents have been safely and sustainably preserved for a long time (multiple backups); metadata are being expertly reconstructed in accordance with international standards (Dublin Core schema); based on priority research criteria, several special collections were selected from the total mass of 1,502 documents (Collection of F. Wollman; Collection of research questionnaires; thematic research collections about viticulture, rural interior design, and textile patchwork); experimental research use of these specialized collections is underway (testing variants of descriptive metadata assignment); ongoing operations are continuously checked and evaluated, and their settings are modified if necessary; in addition to institutional support, diversified financing of some digital curation tasks from project collaborations was secured, while an increase in such collaborations is expected after the launch of the digital repository. These digital curation strategies (including chemical treatment) saved the CRR from devastation, established its sustainable research use and reuse, and preserved it for future generations.

In conclusion, the principle of digital curation is crucial for enriching the preservation and use of a collection's items during the entire life cycle. In the particular case of a research collection such as CRR, however, it is necessary to pay more attention to the increased need for the participation of scholars in the process of digital curation as well as a smaller number of specific users (the user can often only be the researcher himself). These factors determine the digital curation process of research collection and condition the quality of its temporality, inclusiveness, and sustainability.

2.4. Contextual mass

The previous three principles relate to digital collections in general. According to Palmer (2004: 351–352), the next two will be specified for thematic research collections. The principle of contextual mass means a systematical way of collecting resources dependent on the user's co-curation activity, and the user-creator, in this case, means the researcher. Following the previous two principles, it is a personal involvement of the researcher in the creation of the collection through qualified and careful selection of content and research-meaningful management focused on his area of interest: "Whereas in a traditional collection the metadata is chiefly significant to us as a way of locating the specific item we seek, in the digital collection we must attach much greater significance to what Carole Palmer calls 'contextual mass': the density and interconnectedness of the collection with respect to a specific theme or

research agenda, through which the collection can show us patterns that are both relevant and informationally significant” (Flanders, 2014: 163). The purpose is to curate a research collection saturated with rich humanities contextual information. It means that it should be a balance between thematic focus coherence and high density of heterogeneous interconnection (in particular through various types of metadata) to support its research use: “Collections built on a contextual mass model create a system of interrelated sources where different types of materials and different subjects work together to support deep and multifaceted inquiry in an area of research” (Palmer, 2004: 353). The contextual mass is better understood in contrast to the critical mass of large-scale digital libraries, which are heterogeneous in content and dispersed in context (Palmer, Zavalina, Fenlon, 2010). It applies that the more thematically specific and coherent the collection will be, the higher the degree of connectedness and contextuality and the higher potential for use in research.

The contextual mass is more a requirement than a concise methodological principle (especially the meaning of heterogeneous interconnection) and resembles an emergent system. The example of CRR can provide a better understanding. The original analog *Text Archive* had the characteristics of a critical mass because, as a whole, it was thematically heterogeneous (it extensively tried to cover all possible categories of folk culture as demonstrated in the subject system), and the internal contextuality of such a mass was therefore diluted. The reason for that is that in the past, the possibility of applying the principles of curatorship and description when working with physical documents was reduced. New acquisitions were added to CRR mechanically, a description was limited to the assignment of location and subject (often inaccurately and with a low level of granularity), and the document selection for research purposes using the card catalog was laborious.

The situation is quite different in the case of *The Collection of F. Wollman*, extracted from the *Text Archive* and carefully curated by teams of scientific workers. *The Collection of F. Wollman* is the fund of folk tales collected by students of the Slavic Seminar in Bratislava from 1928 to 1947 (Popelková, Zajonc, 2007). It comprises 119 documents and 8.852 text pages that are well-structured and monothematic⁴. Compared to CRR, it has a low thematic and high interconnection density, fulfilling the contextual mass condition. It was not possible only by mechanical extraction from the CRR but by curatorial teamwork, through which additional layers of research-rich data from primary and secondary sources were purposefully added in terms of content and form. This extra work strengthened the coherence and increased the structural connection of the items, thus, the research potential and utilization. If these documents were not curated, they would remain scattered in the thematic mass and sparse connections.

4 In short, during the digital curation of the CRR, two other specialized thematic research collections were selected concerning the project cooperation of the department. They are smaller in scope than The Collection of F. Wollman and focus on viticulture and rural interior design themes. Documents belonging to these collections were expertly selected from CRR, and an experiment is currently underway to improve the granularity and usability of their descriptive metadata.

Finally, this contextual mass generated one of the most significant works of Slovak ethnology, three volumes of Slovak folk fairy tales (*Slovenské ľudové rozprávky* 1, 1993; *Slovenské ľudové rozprávky* 2, 2001; *Slovenské ľudové rozprávky* 3, 2004).

In conclusion, despite the unclear definition, the principle of contextual mass is an essential aspect of the meaningful curation of research collections (in line with the following fifth principle). However, it remains questionable in which way a thematically coherent and internally richly structured collection increases the potential for research use. For example, the specialized collection extracted from CRR (e.g., *Collection of research questionnaires*) has multiple and reasonable interconnections to items from other CRR collections as well as resources such as the Ethnographic Atlas of Slovakia, ETNOFOLK, the Encyclopedia of Slovak Folk Culture, or papers based on data from these questionnaires published in *Slovenský národopis* / *Slovak Ethnology journal*. This example shows that external out-of-collection contextual information can be important for understanding and searching for items in a research collection. It remains open in what way and to what extent the application of interoperability of research collections can support the value of contextual mass.

2.5. Scholarly contribution

The scholarly contribution principle states that the main criterion of thematic research collections is whether they generate added value for the given scientific field during their creation and use. A clear distinction is made here between the two functions of digital resources concerning research. On the one hand, it is a standard supportive function of digital resources that facilitates or speeds up scientific work. On the other hand, it is a specific generative function by which a digital resource as a complex environment productively participates in scholarship. In this sense, Palmer outlines a dynamic model in which there are responsive and feedback relationships between research collection and the research itself. A research collection enables new research, and research, in this case, does not simply consume a digital resource. This reciprocity is the core of the relationship between interpretative heuristics and technologies/infrastructure in digital humanities.

People make science as well as digital collections. Therefore it is natural to perceive the principle of scholarly contribution from the point of view of authorship. Palmer reflects on the issue on two levels: on the one hand, from a quantitative point of view, she points out that the authorship of a digital collection can rarely be attributed to one or a few co-authors. The nature, time, and technological demands of digital collections require collaborative teamwork (2004: 352). This aspect refers to the frequent emphasis on collegiality and connectedness in the digital humanities community (Spiro, 2012; Svensson, 2010). On the other side, from a qualitative point of view, she replaced the traditional concept of the author with the term “creator”. In this way, she follows up on the idea of an institutionally valued figure of a hybrid digital scholar, in which academic subject expertise and technical computing knowledge are combined uniquely (Fraistat, 2012). For the needs of building and

maintaining digital research collections, in addition to knowledge, know-how in content selection and organization are required.

The scholarly contribution principle is naturally applied in the CRR life-cycle as defined by Palmer: “Like other scholarship in the humanities, research takes place in the production of the resource, and research is advanced as a result of it” (2004: 352). In addition to CRR core service (documentation and infrastructural support), this reciprocal model is realized in the cyclic process of creation-use-creation of collection research reports (Figure II.). In short, the ethnologist, as part of the project tasks, carries out the field research in selected localities, where he purposefully and systematically (on the basis of a questionnaire or other prearranged documentation form) obtains raw data on the given topic and prepares photo documentation, which after the completion of the field research is processed and finalized in the form of a standardized research report, after which expert consultations regarding descriptive information, is handed over to the documentation department, where it is assigned an unique ID number, its data and metadata are processed, incorporated into the repository system, or selected and added into specialized collections by the curator, where it is made available for browsing, querying and studying for other interested persons from the academic community, who usually use the results of the given field research to support their theory or to formulate new scientific assumptions or hypotheses that need to be validated/falsified by further empirical field research that generate another new research report and outputs (many projects solved at IESA SAS were inspired by the research of documents from CRR). It is precisely the added value of the scholarly contribution that CRR and the participating actors/factors bring to the ethnological research environment.

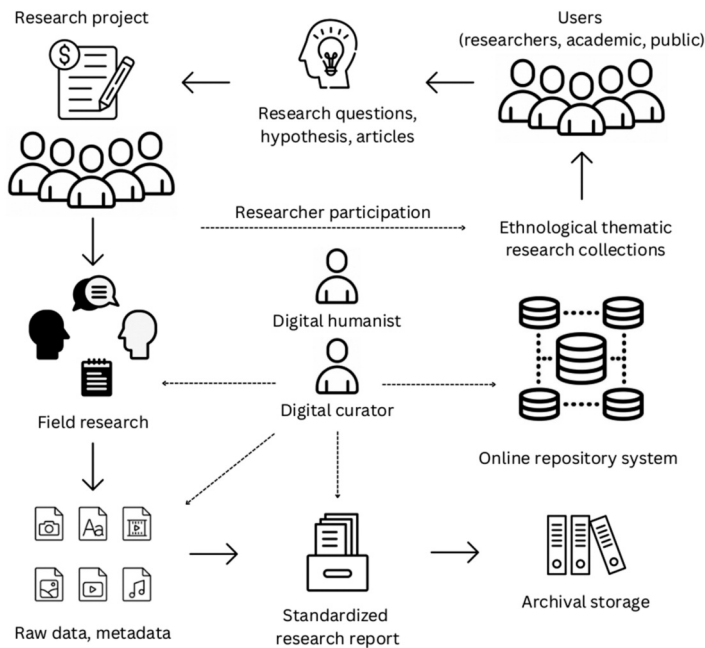


Figure II. CRR digital curation – research environment lifecycle model

Similarly, the issue of authorship has been present in CRR from its beginnings to current digital curation. CRR has been built over decades, during which there have been many personnel and generational exchanges of documentarians with archival knowledge. The total number of authors who have contributed research reports is around 350. In the past, the department had its photographer and drafters, and scientific workers were often part of its staff. An expert in digital humanities devises optimized software solutions, a digital curator plans and manages selection and preservation. Auxiliary personnel with basic technological know-how participated in the bulky processes of digitization and transcriptions. The completion of thematic collections is curated with the participation of experts in the given field, and emeritus and former workers consult the original documentation methods. It is clear that the attempt to assign authorship for both the analog and digital versions of the CRR is questionable. The participating persons-creators show that it is long-term multi-author teamwork with an even distribution of know-how in archiving, documentation, digital curation, digital humanities, ethnology, and applied computer science⁵.

In conclusion, the scholarly contribution principle in the form of the generative function of research collections correctly emphasizes the need for the internal connection of the creation of resources and research activities (so that the research collection is not just a supporting source of information). However, the cyclic process of creation-use-creation for the added value of the research collection needs to be revised from the point of view of its sustainability (working with researchers can be time-consuming; professionally, personally and communication demanding). The request to generate new research and collection contributions means that digital curatorship reaches beyond the controlled collections area into the broader area of the research environment. However, this area is outside the competence of the digital curator because it involves various institutional, personnel, project, and financial variables. So applying this principle often remains at the level of potentiality or even chance.

Summary

In sub-chapters focused on selected principles of research collections, all three specified aims of the article were realized in parallel. First, individual principles were analyzed, then their exemplification through the digital conversion of CRR, and finally, their evaluation. Following the epigraph, which says that there are no absolute rules for creating good digital collections, the paper, through the evaluation of these principles, draws attention to specific modifications of their application for the creation of research collections using digital conversion: the issue of development

5 The department's distribution of know-how necessary for building digital research collections is relatively sufficient. However, in the long term, it is unsustainable, especially considering under-staffing and the absence of a full-time professional programmer.

policy commitment; of descriptive content and format adjustment; of digital curation demand on scholars' participation; of contextual mass external links; and of managing the scholarly contribution research environment potentiality.

Acknowledgements:

The paper was produced with the support of the scientific research project The technological and socio-cultural contexts of innovation in material culture: The ethnological perspective, VEGA no. 2/0051/22.

REFERENCES

- Abbott, D. (2008). "What is Digital Curation?". *DCC Briefing Papers: Introduction to Curation*. Edinburgh: Digital Curation Centre. Handle: 1842/3362. Available online: <https://www.dcc.ac.uk/guidance/briefing-papers/introduction-curation/what-digital-curation>
- Carignan, Y., Evander, J., Gueguen, G., Hanlon, A., Murray, K., Roper, J., Ross, T., Schreibman, S. (2007). *Best Practice Guidelines for Digital Collections at University of Maryland Libraries* (p. 81). Office of Digital Collections and Research University of Maryland. <https://silo.tips/download/best-practice-guidelines-for-digital-collections#modals>
- Cole, T. W., Jackson, A. S., Palmer, C. L., Shreeves, S. L., Twidale, M. B., Zavalina, O. L. (2006). *Findings Pertaining to the Framework of Guidance for Building Good Digital Collections* (p. 13) [White paper]. University of Illinois at Urbana-Champaign. <http://hdl.handle.net/2142/722>
- Duranti, L., Franks, P. C. (Eds.) (2015). *Encyclopedia of archival science*. Rowman & Littlefield.
- Fenlon, K., Senseney, M., Green, H., Bhattacharyya, S., Willis, C., Downie, J. S. (2014). Scholar-built collections: A study of user requirements for research in large-scale digital libraries. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1–10. <https://doi.org/10.1002/meet.2014.14505101047>
- Fenlon, K. (2017). Toward a characterization of digital humanities research collections: A contrastive analysis of technical designs. *Proceedings of the Association for Information Science and Technology*, 54(1), 82–92. <https://doi.org/10.1002/pra2.2017.14505401010>
- Flanders, J. (2014). Rethinking Collections. In: P. L. Arthur, K. Bode (Eds.), *Advancing Digital Humanities* (pp. 163–174). Palgrave Macmillan UK. https://doi.org/10.1057/9781137337016_11
- Fraistat, N. (2012). The Function of Digital Humanities Centers at the Present Time. In: M. K. Gold (Ed.), *Debates in the Digital Humanities* (pp. 281–291). University of Minnesota Press. <https://doi.org/10.5749/minnesota/9780816677948.003.0028>
- Giaretta, D. (2011). *Advanced Digital Preservation*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-16809-3>
- Gill, T. (2008). Metadata and the Web. In: *Introduction to Metadata* (2nd ed., revised by M. Baca). Getty Research Institute.
- Harvey, D. R., Oliver, G. (2016). *Digital curation* (Second edition). ALA Neal-Schuman, an imprint of the American Library Association.
- Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, 3(1), 134–140. <https://doi.org/10.2218/ijdc.v3i1.48>
- Hughes, L., Constantopoulos, P., Dallas, C. (2015). Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines. In: S. Schreibman, R.

- Siemens, J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 150–170). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118680605.ch11>
- Lorenzini, M., Rospocher, M., Tonelli, S. (2021). On assessing metadata completeness in digital cultural heritage repositories. *Digital Scholarship in the Humanities*, 36(Supplement_2), 182–188. <https://doi.org/10.1093/llc/fqab036>
- Miller, A. (2019). Data Visualization as Participatory Research: A Model for Digital Collections to Inspire User-Driven Research. *Journal of Web Librarianship*, 13(2), 127–155. <https://doi.org/10.1080/19322909.2019.1586617>
- NISO (2007). A framework of guidance for building good digital collections: A NISO recommended practice (3rd ed). Baltimore: National Information Standards Organization.
- NSLA (2019). Digital collecting principles. National and State Libraries Australasia. <https://www.nsla.org.au/index.php/resources/digital-collecting-principles>
- Palmer, C. L. (2004). Thematic Research Collections. In: S. Schreibman, R. Siemens, J. Unsworth (Eds.), *A Companion to Digital Humanities* (pp. 348–365). Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470999875.ch24>
- Palmer, C. L., Zavalina, O. L., Fenlon, K. (2010). Beyond Size and Search: Building Contextual Mass in Digital Aggregations for Scholarly Use. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1–10. <https://doi.org/10.1002/meet.14504701213>
- Panczová, Z., Kiliánová, G., Kubisa, T. (2021). *Národopis na Slovensku v službách Tretej ríše*. Bratislava: Ústav etnológie a sociálnej antropológie SAV, Marenčin PT, spol. s r. o. <https://doi.org/10.31577/2021.9788056909591>
- Pennock, M. (2013). The Twelve Principles of Digital Preservation (and a cartridge in a repository...). Collection Care Blog. <https://blogs.bl.uk/collectioncare/2013/09/the-twelve-principles-of-digital-preservation.html>
- Pomerantz, J. (2015). *Metadata*. The MIT Press.
- Popelková, K., Zajonc, J. (2000). Predmetový register. Územný register. In: *Slovensko. Európske kontexty ľudovej kultúry* (pp. 359–380). VEDA.
- Popelková, K., Zajonc, J. (2007). Elektronický katalóg Wollmanovského archívu: Aktuálny stav spracovania. *Národopisná revue*, 17(1), 39–49.
- Prandová, E. (1982). *Dokumentačné fondy Národopisného ústavu SAV 1946–1977*. Krajské osvetové stredisko v Bratislave.
- Sabharwal, A. (2015). *Digital curation in the digital humanities: Preserving and promoting archival and special collections*. Chandos Publishing.
- Slovenské ľudové rozprávky* 1. (1993). Filová, B., Gašparíková, V. (Eds.). Veda.
- Slovenské ľudové rozprávky* 2. (2001). Filová, B., Gašparíková, V. (Eds.). Veda.
- Slovenské ľudové rozprávky* 3. (2004). Filová, B., Gašparíková, V. (Eds.). Veda.
- Spiro, L. (2012). “This Is Why We Fight”: Defining the Values of the Digital Humanities. In: M. K. Gold (Ed.), *Debates in Digital Humanities* (online). University of Minnesota Press. <http://dhdebates.gc.cuny.edu/debates/text/13>
- Svensson, P. (2010). The Landscape of Digital Humanities. *Digital Humanities Quarterly*, 4(1). <http://www.digitalhumanities.org/dhq/vol/4/1/000080/000080.html>
- Unsworth, J. (2000). Thematic Research Collections. MLA Annual Conference, Washington DC. <http://johnunsworth.name/MLA.00/>
- Zajonc, J. (2006). Budovanie elektronických dokumentačných a bibliografických databáz v Ústave etnológie SAV. *Etnologické rozpravy*, 13(1), 30–47.
- Žatko, R. (1953). *Vecný systém triedenia materiálu*. Národopisný ústav SAV.

ABOUT THE AUTHOR

ANDREJ GOGORA (ORCID: 0000-0001-5169-5330) – is a researcher at the Institute of Ethnology and Social Anthropology SAS and Head of the Research Collections Department. He completed his doctoral studies in systematic philosophy at the Institute of Philosophy SAS. He is engaged in basic research on the methodological foundations of digital humanities and digital curation. In applied research, he is involved in building ethnological digital research collections. He is the coordinator on behalf of the IESA SAS as a cooperating partner of the European research infrastructure DARIAH-EU.