Christoph Schmitt, Alf-Christian Schering

# WossiDiA

## A Pioneer of Digitized Tradition Archives and Its Hypergraph Technology

https://doi.org/10.1515/fabula-2023-0007

**Abstract:** This article sheds light on the digital transformation of Richard Wossidlo's regional ethnographic and linguistic collection into the research environment WossiDiA. The Digital Wossidlo Archive is based on making handwritten sources and their indexing visible to users. It has been implemented as a hypergraph database, whose abundance of node and hyperedge types is suitable for representing the diversity and delicate nature of the tradition archive. This paper will discuss how proprietary properties have been preserved in the transformation process and how the digital collection is compatible with big data scenarios. Follow-up projects have contributed to this. "Efficiency," the necessity of resource-saving, will also be problematized. Finally, it will be shown how a sparse ethnographic fieldnote can be enriched into a complex hypergraph extract, a method which is reminiscent of dragnet detection.

**Zusammenfassung:** Dieser Beitrag beleuchtet die digitale Transformation der regionalethnografischen und regionalsprachlichen Sammlung Richard Wossidlos zur Forschungsumgebung WossiDiA. Das digitale Wossidlo-Archiv dient dazu, Nutzerinnen und Nutzern handschriftliche Quellen und deren Erschließung zugänglich zu machen. Es wurde als Hypergraph-Datenbank realisiert, deren Fülle an Knoten- und Hyperkantentypen sich dazu eignet, die Vielfältigkeit und spezielle Beschaffenheit des Traditionsarchivs abzubilden. Diskutiert wird, wie die Eigentümlichkeiten von Wossidlos Sammlung im digitalen Transformationsprozess erhalten wurden und die Sammlung gleichwohl für Big-Data-Szenarien anschlussfähig gemacht werden konnte, wozu auch Folgeprojekte beitrugen. Unter dem Stichwort „Effizienz" wird die Notwendigkeit, Ressourcen zu sparen, erörtert. Schließlich wird gezeigt, wie aus einer knappen ethnografischen Feldnotiz ein komplexer Hypergraph-Auszug werden kann – ein Verfahren, das an Rasterfahndung erinnert.

**Alf-Christian Schering**, Research assistant at the Database Research Group at the University of Rostock (e.g. WossiDiA project and its follow-up projects) and at the Serbski institut/Sorbian Institute in Bautzen in the project "Register for the Sorbian Cultural Heritage". E-Mail: ac@informatik.uni-rostock.de
**Christoph Schmitt**, until 2022 Head of the Wossidlo Research Center for European Ethnology/Folklore at the University of Rostock. E-Mail: christoph.schmitt@uni-rostock.de

# 1 Introduction

In 2014, the first version of WossiDiA, the Digital Wossidlo Archive, was published online. Using a hypergraph database, WossiDiA displays the folklore collection of Richard Wossidlo (1859–1939). The grammar schoolteacher and private ethnographic scholar carried out his survey over fifty years in the Mecklenburg area. His estate also documents the lexis and use of the dialects spoken here until the outbreak of the Second World War. Wossidlo's collection has highly proprietary features, consisting of a sophisticated card box system and associated corpora. Hypergraph technology is a novelty in the field of archives and collections. As will be shown, it has succeeded in bringing the idiosyncratic structure of Wossidlo's empirical data collection closer to today's users. And since nearly all of its archival sources have been published via WossiDiA (that is, not only finding aids and samples but two million scans of mostly handwritten documents), we count ourselves among the pioneers in the field of digitally transformed folklore archives.

Reason enough to reflect on the problems to overcome and point out the added value of the WossiDiA project. These added values are mainly based on innovative technologies. Folklorists are challenged to grasp the potential of state-of-the-art techniques to some extent. This requires intensive cooperation between the humanities and information technology disciplines. Computative techniques accelerate the work process and offer innovative navigation and data mining options. In particular, graphs – especially their potentiation to so-called hypergraphs – can generate different search paths and views to provoke new questions. It should also be noted that computer-aided methods can be used to research questions that previously seemed too time-consuming to answer.

At the beginning of this article, the visibility of Wossidlo's collection will be discussed, which has been increased due to its online placement and two consecutive projects: "Intelligent Search Engine for Belief Legends (ISEBEL)"[1] (that is discussed in this volume) and the current project 'Wossidlo-Teuchert' online. Publication of the Mecklenburg dictionary in the Trier dictionary network and corpus-based bidirectional linking with the 'WossiDiA' digital research environment.[2] Then the complexity of the folklore archive, its abundance of traditional forms of expression, topics, content, metadata, and cross-references will be explained to show how hypergraphs have modeled this diversity of relationships.

The reach of WossiDiA has been significantly improved by the construction of new interfaces, which will be briefly outlined here. Under the heading of efficiency,

1 http://search.isebel.eu (March 10, 2023).

2 https://tcdh.uni-trier.de/de/projekt/wossidlo-teuchert-online (March 10, 2023).

optimizations of the system will be discussed. Optimizing collection activities for time reasons was already a concern of the older ethnographers. In the last part, the networking logic of Wossidlo's card box system, which was made to speak with the help of the hypergraph database, will be demonstrated. An example is a "thin" fieldnote about a witch legend which will be enriched by connecting it to neighboring sources, related genres, and different cultural forms of expression (rites, thing culture, etc.) and the linguistic system. Therefore, it is by aggregation that a semantic analysis is made possible. A large number of hyperedges and nodes makes it possible to complete scattered information, which is reminiscent of dragnet search. The visualization of graph-mining scenarios of the data basis available in WossiDiA, e.g., with the query language GrafL (cf. Meyer/Mukbhil/Schering 2017), will not be examined.

## 2 From black boxes to open archives

Of course, judgments about the Wossidlo collection depend on knowledge of the unique archive. After Kaarle Krohn had visited the collection of Wossidlo in Waren an der Müritz, he advertised Wossidlo's enthusiastic commitment in the first communication of the Folklore Fellows. He named the collections of the Danish teacher Evald Tang Kristensen and the Estonian pastor Jacob Hurt as further models (Krohn 1907, 1). In 1925, Archer Taylor similarly combined his trip to Europe with a visit to these three collections (Schmitt 2015, 25). Gottfried Henßen, who drew from Wossidlo's collection on site from 1936 onwards in two consecutive winter months, commented retrospectively: "What I found in Waren was downright overwhelming" (Henßen 1959, IX). At this time, Henßen was busy setting up what would later become the Zentralarchiv der deutschen Volkserzählung (Becker 2005).

For most researchers, however, the collection – accelerated by the division of Germany – had to appear like a black box. For example, Will-Erich Peuckert was disturbed by Wossidlo's presentation of legends in his *Mecklenburgische Sagen*, which seemed to Peuckert like a "register"; even motif research would not be able to do anything with it (Peuckert/Lauffer 1951, 193). But a completely different picture emerged in the 1960s when Lutz Röhrich began working with the Wossidlo-Forschungsstelle (which was opened in Rostock in 1954). Wossidlo's densely documented legends and indexing method contributed significantly to the creation of Röhrich's catalog of legends about death and the dead (Röhrich 1967). Peuckert had never inspected the collection, while Röhrich received more information about it through the mediation of Ingeborg Müller, who worked in the Wossidlo Archive at that time.

In German-speaking European Ethnology/Cultural Anthropology/Folklore Wossidlo was rediscovered at the end of the 1970s and made the ancestor of modern field research, which rejected the previous "Gewährsleute"-principle and considered dialogue with the informants necessary (Göttsch-Elten 2009, 18). One more reason to continue working with his collection.

The black box that the archive largely represented has now turned transparent. More precisely, 1,819,708 scanned pages of the Wossidlo estate, including the oral sources for the dictionary of the Mecklenburg dialect (Wossidlo/Teuchert 1942–1992), can currently be viewed. Although users can illuminate the interior of the information space at will, seeing correctly and navigating excellently requires a basic understanding of the collection and the digital archive system. Since no "real" archivist can help, the digital archive has to explain itself. Carrying out this mediation work is a central concern of WossiDiA. This is not easy because a complete collection is now "exposed," not only in selection by printed editions. Let us, therefore, take a closer look at the previous editing styles with which parts of the Wossidlo collection were published. Then it will be easier to assess what the digital archive can and cannot do.

Printed editions are usually a selection determined by research interests, personal preferences, shared tastes, self-censorship, or institutional constraints. Stories must be tellable, which requires slight interventions. Siegfried Neumann edited, among others, fairy tales (Neumann 1971) and droll stories (Neumann 1963) from Wossidlo's collection. Although his editions include Wossidlo's keyword-like field notes, they are primarily based on the well-told stories of Wossidlo's corresponding collection helpers. Neumann's adaptations include explanations of dialectal lexis. In this way, he conveyed the Low German texts beyond the lay audience to folk narrative research, to which his indexing work and numerous articles for the *Enzyklopädie des Märchens* contributed.

The fairy tales and legends indexed and edited by Neumann contain references to additional (unpublished) sources of the Wossidlo Archive. If users enter the relevant tale type in the search field of WossiDiA, both the published and the unpublished manuscripts are shown (if these entries have yet been made, which only happens gradually). This way, they can compare the original version with the edition text and view the unknown versions of a topic. In other words, users can retrace selection processes and text interventions.

Field notes made by Wossidlo are more difficult to convey. He wrote down the fruits of his fieldwork in fleeting handwriting in pencil. After a direct encounter in the field, he memorized numerous details, keywords, and phrases, which he recorded as soon as possible (cf. Emerson/Fretz/Shaw 2011, 31). He then added to his still-fresh memories on more comfortable occasions. Since Wossidlo used numerous abbreviations, these had to be decoded by WossiDiA. His sparse fieldnotes become

entirely understandable through their position in the hierarchically ordered and related categories. Their place in classified bundles, card boxes, and rows of card boxes is via WossiDiA much better visible than on site at the "real" archive. Against this background, Peuckert's judgment mentioned above makes more sense. Wossidlo hardly presented complete narratives but arranged motif variants according to the morphology of large groups of legends (such as witch legends, devil legends, etc.). For a more thorough use of the legends, one needs the scans of the original documents beyond the printed version of the *Mecklenburgische Sagen*.

Gisela Schneidewind proved in her edition of the sacrilegious legends of cruel nobles (Schneidewind 1960) the validity of Wossidlo's sparse fieldnotes. It grows when variants are placed next to each other and recording metadata are supplied. This was done using a "narrator register" that lists additional metadata extracted from Wossidlo's diary notes. Also listed are the narrators' professions (or employments), their place of birth, age, relationship with other narrators, and Wossidlo's comments on their repertoire. The core of a legend and its variability – no embellished version, but the tradition itself – becomes known. Schneidewind's edition has a more documentary character, although the thematic selection ("master and servant") was determined by the paradigm of GDR folklore, which was still young at that time.

With the help of WossiDiA, fragmentary or more complete documents of the collection can be called up within seconds via various accesses and displayed in the hypergraph network. Nevertheless, further development of the virtual archive and annotations are desirable. Due to the abundance of sources, the virtual archive cannot provide the mediation and explanatory power of classical editing methods across the board but gradually through in-depth indexing of selected collection areas. Instead, WossiDiA offers almost inexhaustible knowledge stored in the network hypergraph to get hold of it in the right place. The system can represent the development of volatile fieldnotes through intermediate stages (such as fair copies or transcriptions) to printing. Relevant excerpts from printed sources can be viewed, and users can be linked to the Network of Digital Libraries.

Many other scenarios are conceivable. For example, correspondence with informants or scholars about certain topics or the developing scientific community can be accurately located in the WossiDiA thesaurus. The data collected on a particular day can be bundled to reconstruct field research trips. Analyzed diary notes can contribute to this. Some of these scenarios have been implemented at least rudimentarily.

In their early days, many collections were not intended for the public but were designed as personal aids. Even though Wossidlo's collection areas later expanded, the rules he set for writing ethnographic fieldnotes and indexing can be recognized at the beginning of his fieldwork (Schmitt 2019, 142). Therefore, they mostly stayed

the same in more than fifty years of collecting. In the database, each document can be accessed via a keyword assigned by Wossidlo. To quite a few of them, we have given modern terms. In the thesaurus, one can navigate through the hierarchical structure or across other expressive forms (narratives, customs, superstition, object culture, semantic fields of language, etc.).

After putting the WossiDiA prototype online, many more handwritten documents were made computatively readable. Step by step, more keywords and metadata were entered, which is a never-ending process. Typewritten transcriptions on paper forms from earlier times could be read in with OCR but had to be corrected manually. The use of computational handwriting recognition has not made sense to us, given the enormous number of writers' hands.

## 3 Modeling of relationship diversity by using hypergraphs

Wossidlo set himself the task of grasping the oral tradition in the diversity of its cultural and linguistic forms of expression. The collection was limited only by the scope of the survey, i.e., by the Mecklenburgian area. The slip-of-paper technique is symptomatic of the vision of a holistic recording (Dietzsch 2009, 23). Wossidlo was a philologist and a "friend of the word" (Henßen 1957, VII). Because he also documented the language used, his collection extends far beyond the classical folkloristic canon. And since his collection enterprise includes various survey methods – from direct exploration, the correspondent procedure to questionnaires – it results in a vast network of new empirical data enriched by excerpts from written sources. Because these data are interwoven, it was necessary to map this network with the help of hypergraphs, whose technique will not be explained in more detail here (see Meyer/Schering/Schmitt 2014 and Meyer/Schering/Heuer 2017).

Only this much: Graphs are abstract structures that can be imagined analogously to sentence-like relationship structures. Subjects and objects are represented by nodes to be specified in more detail, and the relationships between them, comparable to predicates, by so-called edges. Graphs can be determined in more fact by typifications, attributions, and directions. Hypergraphs are superimposed graphs with common nodes and edges being connected and, therefore, navigable.

Let's imagine this: There is a legend circulating that takes place in a specific location. A person tells this legend, and another person records the story on behalf of a third person (= W.) and sends him the story. W. tags the story with some keywords using a set of abbreviations. He notes the place of action on top of the slip of

paper. Now W. puts this slip of paper in a bundle, whose legends entwine around precisely this location. Since W. not only has a local chronicle interest but wants to research the supra-regional tradition, he duplicates part of the slip of paper and puts it in a suitable group of legends.

Later, W. meets a woman who happens to tell him an almost similar legend. Since W. already knows the story, he limits himself to note deviations. The legend also contains a proverb, which W. records and later duplicates on another slip of paper. Since the female narrator knows even more stories, W. captures them in keywords. In one of the stories, a significant requisite is an essential part of a custom. Later, he will mark this object as a motif and link it to the legend and the type of custom. In the evening, W. again looks through his diary notes and records what he knows about the narrator. How old was she, and how did she and her husband earn their living? Once home, W. looks through his previous notes again, including his excerpts from printed versions. Will he find the variant confirmed by similar ones or will it be different? Now he notices that he could have asked a lot better. Since this concerns several cases, he formulates a few specific questions, which he sends to some of his collection helpers and publishes in a regional magazine.

By using hypergraphs such filigree and complex scenarios can be mapped if enough nodes and edges have been defined. On the one hand, persons, places, stories, literature, words, documents, tale types, motifs, etc., were defined as node types and classified in more detail with attributes, i.e., persons with name and gender, etc. On the other hand, the relevant relationships between them – i.e., whether people communicate orally or in writing, whether places are part of the story content or recording data, and whether words function as keywords or synonyms – were described with hyperedges.

The WossiDiA-PowerGraph catalog currently has 74 node types and 115 hyperedge types. One can imagine how complex the WossiDiA hypergraph can be fed, so that deep drilling is possible across all scenarios. The WossiDiA hypergraph currently contains 524,195 relationships involving 1,138,279 nodes. Typically, two nodes are linked to a hyperedge. But this is not a restriction in hypergraphs.

Depending on the hyperedge-type model, more or fewer nodes can be involved. In WossiDiA, for example, the hyperedge type “origin” only connects people and their place of origin. In contrast, the hyperedge type “narrated_oral” can link people, their professions, the place of narration, the recording time, and other nodes. Some entries are mandatory, such as the keywords or word entries in the headers of the slips of paper; others are optional due to time constraints.

Table 1 shows statistics of the most commonly used node and hyperedge types. It represents only a snapshot of the description process, so it does not serve as a basis for an exhaustive analysis of the corpora involved. However, tendencies can

already be discerned. With nearly 300,000 occurrences, "word nodes" (type ID 18) are the most used, accounting for a quarter of the 1,1 million entries of nodes. This shows the extent and depth of Wossidlo's keywording, including word fields with synonyms and the Low German lexis. The number of different (node-based) words entered so far (including keywords) is currently 54,026 and is increasing rapidly due to the now-running dictionary project. It also becomes apparent that the collector retains the word-based approach across all forms of folk culture. The distinct functions assigned to this "word node" are discussed below.

The node "ZAW 2" (ID 103) was developed for the systematically ordered collection of slips of paper. It is based on 920,747 slips of paper, classified by 29,713 labeled bundles, compiled into 1,091 boxes sorted according to several types of tradition. This node has now been linked 150,000 times, which shows that many entries are still needed here. About sixty percent of these documents represent excerpts from folkloristic, ethnographic, and linguistic publications. Their abbreviations have currently been entered rudimentarily (ID 13 = 67,494). About forty percent are Wossidlo's field notes and links to keywords from correspondences. Although the jottings of his fieldwork, written on single slips of paper, are provided with a primary keyword, the recording data noted by him (informant, place of recording, informant's place of origin, time of recording) can only be entered gradually into WossiDiA because this work has to be done manually.

**Tab. 1:** Top-20-statistics of node types participating in hyperedges (left) and hyperedge types (right)

| ID | Node type name | count | ID | Hyperedge type name | count |
|---|---|---|---|---|---|
| **18** | word | 297815 | 1019 | Content | 111954 |
| **103** | ZAW-2: single note | 155560 | 7 | reference (undirected) | 56189 |
| **17** | transcription | 86230 | 1034 | Mecklenburgian Dictionary describing | 47796 |
| **13** | abbreviation of literature | 67494 | 1025 | language-specific | 46284 |
| **133** | MWT-3: dictionary catalog Teuchert (page) | 59097 | 342 | transcription (WossiDiA) | 38663 |
| **16** | translation | 47317 | 8 | reference (directed) | 34041 |
| **102** | ZAW-1: a categorized bundle of notes | 37211 | 302 | abbreviation of literature | 32405 |
| **113** | BKW-3: correspondence with an informant (page) | 35138 | 90 | sem: hierarchy (for thesaurus) | 17522 |
| **142** | ZTW-2: transcription (document) | 31784 | 70 | narrated | 16156 |
| **99** | date/duration | 31018 | 71 | narrated (orally) | 15465 |
| **27** | field name | 24203 | 80 | synonym | 15454 |
| **122** | MWW-2: dictionary catalog Wossidlo (subbundle) | 19101 | 1028 | story transcription | 14274 |
| **22** | document type | 18408 | 62 | questionnaire answer | 12063 |

**Tab. 1:** (continued)

| ID | Node type name | count | ID | Hyperedge type name | count |
|---|---|---|---|---|---|
| **124** | MWW-4: dictionary catalog Wossidlo (page) | 18289 | 91 | Sem: reference (for thesaurus) | 7473 |
| **14** | abbreviation of word | 15575 | 301 | abbreviation of word | 6317 |
| **19** | Wossidlo questionnaire | 14748 | 72 | narrated in writing | 5708 |
| **29** | ISEBEL story | 14364 | 81 | activity | 5308 |
| **111** | BKW-1: contribution | 13254 | 1020 | document type | 4598 |
| **112** | BKW-2: correspondence with an informant (part) | 9283 | 1010 | affiliation/role/place | 3661 |
| **123** | MWW-3: sub-subbundle | 5663 | 341 | transcription (academy) | 3100 |

The letters of his helpers doing fieldwork ("BKW") are tagged by keywords, which Wossidlo duplicated on slips of paper on which the page of the letter in question is noted. Unfortunately, these links also have to be entered manually; more than 20,000 (= ID No. 111 and 112) have already been set. The systematically ordered slips of paper consist of 981,062 pages, which shows that most of them (such as literary excerpts) are one-sided. Fieldnotes are one- or two-sided and rarely contain multiple pages. The third most frequently used node captures transcriptions of documents (ID 17). About 30,000 transcription texts could be taken over, mainly from paper forms, and a minor part was newly created. This stock can also be searched via text recognition. The language node type (ID 16) has been used nearly 50,000 times and includes Low German and English words. Collection-specific indexing, i.e., Wossidlo's keywording, was mainly translated into English and thus made internationally compatible. For WossiDiA's participation in the international harvester Intelligent Search Engine for Belief Legends (ISEBEL), a node type for stories (ID 29) has been developed. Currently, 14,235 stories are offered for ISEBEL.

It is noteworthy that the word abbreviation node type (ID 14) has been used over 14,000 times. Abbreviations of literature are not included here. The number of different word abbreviations is currently 1,043. This illustrates how much Wossidlo sought to speed up his data collection and management process. Assignments by tale-type indices and motif registers have presently been developed on a trial basis, as ID 26 shows.

The most commonly used hyperedge type is ID 1019, which relates content to document nodes. So far, 111,954 hyperedges have been equipped with 240,790 nodes. ID 18 shows that the number of "word node" usages is higher (= 297,815) because certain words are used to fill other hyperedge types: "synonym" (ID 80) with currently 15,454 uses and hyperedge types marking hierarchical relationships (ID 90 = 17,522 uses), cross-references (ID 91 = 7,493 uses), and synonymous terms

(ID 93 = 2,654 uses) within the WossiDiA thesaurus. The rest is distributed among other hyperedge types.

The hyperedge type “content” (ID 1019) consists of three “LinkRoles”: With the first role, source groups are linked, with the second content, and with the third contexts to the content in question. For context-related entries, the node type ID 18 (“word node”) can be used. They are rare, not least because they are assigned intuitively (1,431 uses = one percent of ID 1019). Another hyperedge bundle will explain the interaction of nodes and their relationships shaped by “LinkRoles”: To distinguish oral from written communication, the two hyperedges “narrated orally” (ID 71) and “narrated in writing” (ID 72) have been developed. If the message medium has to remain open due to missing data, hyperedge ID 73 (“narrated”) is used.

ID 71 was filled 15,465 times, and ID 72 5,708 times. While the hyperedge type “narrated orally” broadly captures Wossidlo’s field research notes, ID 72 mainly marks links to the correspondence of his helpers doing fieldwork. Both hyperedge types can be specified with six different “LinkRoles” (LR). LR 72.1 links nodes of the archive topology, LR 72.2 the informants and their professions, LR 72.3 the place of recording, LR 72.4 the informant’s place of origin, and LR 72.5 the time of recording, while LR 72.6 marks informants of the informant. The latter has only been filled in one percent of the hyperedge usage but would be much higher in the case of in-depth development of the correspondence. Using LR 72.3 and LR 72.4, internal migrations of the informants can be determined.

According to their first impression, many users of WossiDiA are drowning in fragmented documents. However, its high degree of granulation is represented with correspondingly networked structures and their functions, which can be navigated by facilized search. The advantage of hyperedges is that they can be set dynamically. To solve unique problems, tailored node and hyperedge types can be implemented without much effort.

## 4 Interfaces for texts and techniques

To communicate with other digital knowledge databases, suitable interfaces are required. In the analog world of printed products, publications are compatible due to library standards. Tables of content, indexes of places, persons, and subjects enable faster access; translations mediate between different language areas. Folk narrative research is based on international indices; specific genres can be communicated worldwide through tale-type registers. Narrative minimal elements can be coded language independently and across genres using an international motif

index (MOT, Thompson 1955–1958). Traditional legends are a focus of many folklore archives. This narrative genre forms micro-traditions that make supra-regional cataloging more difficult (Schmitt/Tangherlini 2019, 181–182). In the past, therefore, many regional legends were hardly noticed across borders. A novelty – not only from a database point of view – is the harvester ISEBEL, described in more detail in this anthology[3]. It is a common interface offering exceptional geo-based views and navigation forms and enables automatic methods for the comparative search of belief legends. The harvester allows to travel there and back to the peripheral archives, where the records can be tracked more closely.

Several attributes have been assigned to the node "tale type" (ID 26). This allows to build references to ATU (Uther 2004) and national or regional tale-type catalogs. Internal signatures of texts from other folklore archives than the partners of ISEBEL can also be linked with this node type (see the entries of some numbers of the Zentralarchiv der deutschen Volkserzählung in Marburg, where copies of the Wossidlo collection once found their way).

A further interface will be created for the current project: 'Wossidlo-Teuchert' online. It aims to link WossiDiA with the Mecklenburg dictionary (*Wossidlo-Teuchert Mecklenburgisches Wörterbuch* 1942–1992) after it has been digitized and marked up by XML. This will turn the sources from which the dictionary has been generated widely visible. WossiDiA will thus be compatible beyond ethnography for dialectology. As a result, alphabetically ordered card documents (= 778,936 documents) and the above mentioned systematically structured slips of paper (= 920,747 records) will be linked to the dictionary first and foremost via a node type for lemmata. The project will create a multi-perspective approach, for example, via word fields or a High German-Low German reverse lexicon. Its added value will also be profitable for folk narrative research. Because dealing with words and phrases of Low German narrative texts will become easier. Subtle translations from (northeastern) Low German into high-standard German can be done. And on this basis, automatically supported translations into foreign languages can be carried out.

In addition, the international connectivity underlying the long-term project *Enzyklopädie des Märchens* must be exploited. The thesaurus-like conception of the index presented in the final volume (*Enzyklopädie des Märchens* 2015, 223–1115) is in this regard instrumental. At the top level, keywords are grouped by content groups into more abstract keywords. On the lower level, "narrative elements" are recorded as fine-grained units, and larger units of meaning follow a main lemma. The tale types and motifs are also listed. However, the online search engine created

3 See the article by Meder/Himstedt-Vaid/Meyer in this volume.

by de Gruyter Verlag is not nearly suitable for exploiting the computational possibilities described here. It would therefore make sense to include this register in big data scenarios of folklore archives (Schmitt 2022, 135).

Of course, this also applies to working with type indices. In the ISEBEL project, which achieves connectivity via topic modeling and automatic translation procedures, the value of the collector-specific indices became apparent, which were therefore integrated into the harvester. Because these are abstract categories, they cannot be created by automatic keywording. The work on the Irish National Folklore Collection (Ryan/Cárthaigh 2018) shows how the thesauri of older collections can be made compatible with the controlled vocabulary of digital libraries and the language use of today's thesauri. Using the fairy tales collected by Evald Tang Kristensen, the research group led by Tim Tangherlini demonstrates how Kristensen's index, the ATU index, and a later indexing work of Kristensen's fairy tales can be included with the help of network graphs.[4]

As part of the WossiDiA project, two interfaces were initially created: an input interface (WossiApp), reserved for use by the archive staff, and a presentation interface (WossiWeb), which makes the digital content of the archive available to a broad public. The consecutive projects based on WossiDiA required system extensions and additional interfaces. As a result, WossiDiA was transferred from a monolithic software block into a modularized layer-based architecture. In the first step, the application logic of the ethnographic knowledge system was detached from the components and data structures of the hypergraph database system PowerGraph (cf. Schering 2021: 141 and Fig. p. 142). This separation allows to build the logic of other applications based on PowerGraph, whereby hypergraph's basic functionalities are shared and do not have to be developed repeatedly for further projects. In 2019, the WossiDiAPowerGraph server was finally created, which now forms the only contact point for individual applications. A secure Transport Layer Security-based (TLS-based) interface is used for communication between the WossiDiAPowerGraph server and distinctive applications.

Based on this modularized structure, a Representational State Transfer (REST) interface was created for automated access to the WossiDiA hypergraph and to facilitate communication with web-based applications. Furthermore, an Open Archives Initiative (OAI) provider was implemented, which generates complex documents aggregated from the hypergraph for the ISEBEL project, to feed the interface with story data used in this context. To sustainably secure the digital archive via the infrastructure provided by Rostock University Library, an exporter has been

**4** See the article by James Abello et al. in this volume.

developed based on the METS (Metadata Encoding and Transmission Format) and MODS (Metadata Object Description Schema) standards, which are widely used in the library environment.

## 5 Efficiency: Optimization of input, data processing, and storage

Mining algorithms can only be successful if sources are provided with as much metadata as possible. Looking at "efficiency" another aspect comes into focus. Wossidlo already saw the need to organize his ethnographic fieldwork to save time and space (Schmitt 2015: 36–44). Although in WossiDiA the digitization of the sources was delegated to an external service provider, the data transferred to the system had to be processed manually. For this purpose, it was necessary to develop highly specialized software which could provide fast input and efficient description. This resulted in the above-mentioned "WossiApp," a standalone online application with project-internal access.

The increasing complexity of the WossiDiA hypergraph led to data processing that took more and more time. Users had to wait longer and longer for their research results. This was particularly the case with general and highly interconnected nodes, such as terms like "birth" or "marriage". While the individual nodes in the relational backend of the Hypergraph database system were already effectively managed, efficient access to the hyperedges was no longer sufficiently guaranteed as the network became more interconnected. Various index types have been implemented to remedy this condition and equip the system to process even more significant amounts of data, including the Node Abstract Texts (NAT) index, which indexes hypergraph-based text representations of nodes (cf. Schering 2021: 144). Node-to-Node Path Indexes (N2N) were developed to "shorten" typically used search paths and the necessary node and edge constellations or to free them from unnecessary node and edge ballast. They are needed to construct templates, e.g., for a map-based overview of places or the story viewer developed in the context of the ISEBEL project. The latter allows transcriptions, handwritten source documents, and all associated metadata to be presented without significant delay.

Templates that improve usability were developed after the launch of WossiDiA (2014). In addition to node-based list views for metadata in the Web application, the hyperedges connecting the nodes are displayed now when browsing along horizontal and vertical structures of the collection. This allows users to move on from node to node. Later, filters were created for the selection of hyperedge types. Generic

node view templates for archive topology and simple node attribute listings have been retained.

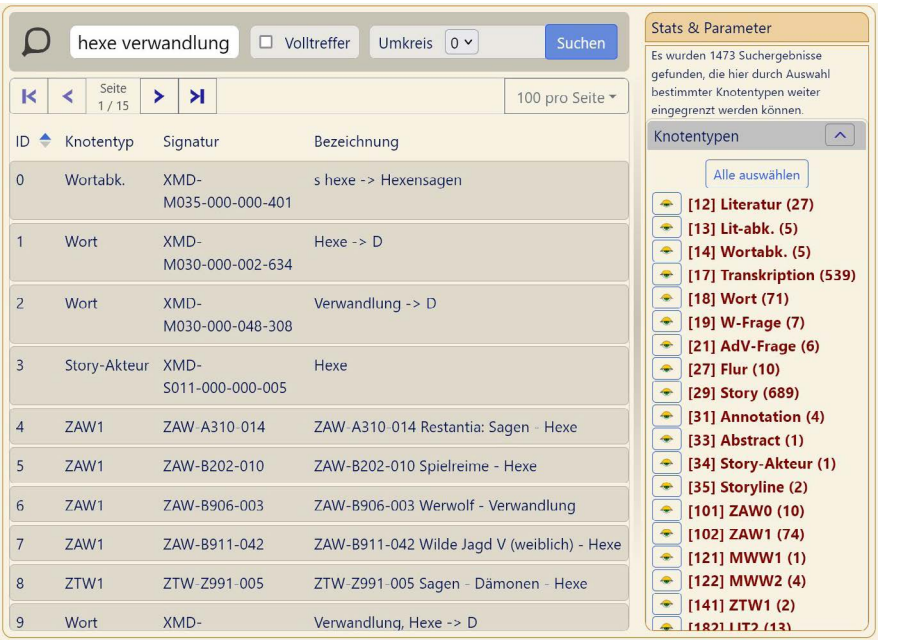



**Fig. 1:** Hits after entering “hexe verwandlung” (witch transformation) into the search field of WossiDiA, filtered by node types

The keyword-based search results (e.g., “hexe verwandlung”) are returned as node lists. Search yields can be faceted by the node types involved (see Fig. 1). If you click on one of the node types, e.g., “Transkription” (ID 17), the search results are filtered accordingly. And if you then click on one of the listed documents, you get to its landing page in the generic view. Templates try to present a node according to the unique characteristics of its type. An example is the person template, which shows a compact presentation of biographical information extracted from the hypergraph, along with the documents provided by that person (including handwritten and transcribed narrations).

# 6 Puzzle work for semantic analysis: Aggregating jottings of ethnographic fieldnotes

Finally, using the example of a fieldnote with sparse jottings, it will be shown how the content and metadata of a single slip of paper can be enriched by its networking within the WossiDiA hypergraph.[5] The aim is to demonstrate the logic and potency of the hypergraph in bringing together cultural and linguistic aspects, which may evoke research questions. A user wants to know into which animals a witch can turn. Via the search field or the thesaurus, he gets to the morphological bundles "witch as animal" (ZAW-B703-001), "three-legged hare" (ZAW-B703-002), "cat, rat, etc." (ZAW-B703-003) and "birds" (ZAW-B703-004). The whole box contains 497 documents. The transformation motif of the "three-legged hare" (cf. *Mot. D 655.2* and *Mot. G 211.2.7*) forms the majority with 70 percent (= 352 notes). So, this motif was widespread in Mecklenburg.

Among the "bird figures" into which witches turn are crows (ZAW-B703-003-001) and black hens (ZAW-B703-004-009). They are well-known as demonic animals. Browsing more documents of the category, the user will meet the following fieldnote: "Rambow/Bihrensch macht sich zu bollnoorst huhn und fliegt nach Rothenmoor/F Löhtz Stöckensee" (*Rambow/Bihrensch transforms herself into bollnoorst chicken and flies to Rothenmoor/Mrs. Löhtz Stöckensee*) (ZAW-B703-004-005).

What is the meaning of the Low German word "bollnoorst"? Does the hypergraph of WossiDiA provide information about the curious chicken, the two places of action (Rabenow and Rothenmoor), and the place of recording (Stöckensee)?

In the right column of the generic view, in which every single document is embedded, the connected hyperedges are shown (see Fig. 2 depicting the fieldnote concerned). The link list of hyperedges is structured according to the logic of the communication process. At the very top, the hyperedge "narrated orally" (edge type ID 71) provides information about the person (node type ID 10). It is a woman named Löhtz. If you click on the person node connected to the hyperedge "narrated orally" (edge type ID 71), you end up on the person template mentioned above. Here the repertoire of the informant becomes visible. It contains more witch legends including a variant of the transformation motif "three-legged hare" (ZAW-B703-003-197).

With the hyperedge type "content" (ID 1019), the compressed content is assigned to the narrative process via keywords ("chicken," "flying," the places of action, and the Low German adjective "bollnoorst"). A click on the latter word, brings one to

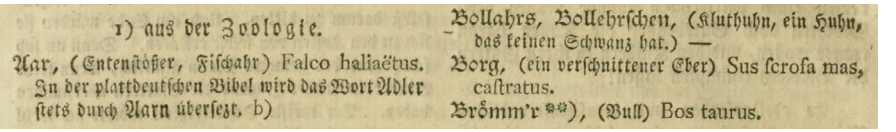
5 Readers can re-enact the following scenario by entering the respective signatures into the search field of WossiDiA.


**Fig. 2:** Generic view of a curious field note (ZAW-B703-004-005)

Wossidlo's alphabetically ordered excerpts for the Mecklenburg dictionary. On some of the slips of paper the "bollnoorst" chicken is explained as a tailless breed. One document refers to the transformation motif by witchcraft (MWW-B011-009-083-013). The sub-subset of the convolute concerned contains several oral references to the noun "Bollnoors," form variants such as "Bollmoors" and the derived adjective "bollnoorst." In the overlying subset "Boll …" (MWW-B011-009-083), composites with the determinant "Boll" (which mainly refers to round objects) are collected (MWW-N061-023-003).

Wossidlo primarily marked the communicative context of language use and sociolinguistic references (i.e., the diaphasic and diastratic levels) but also noted the place of language use (i.e., the diatopic reference). So, he noticed that dogs born without tails were also called "bollnoorst" (MWW-B011-009-083-025). The chicken breed was native but was perceived as exceptional. For example, the adjective "bollnoorst" was used when someone looked strange (MWW-B011-009-083-009) or walked strangely, for which more detailed examples are given: e.g., when someone wore a hat without a brim (MWW-B011-009-083-011) or – in misogynistic meaning – the gait of a woman was associated with the strange chicken (MWW-B011-009-083-029). However, sailors in short vests, whose muscular buttocks stand out more clearly, could also be called this way (MWW-B011-009-083-031).

Users who want to learn more about the gait, appearance, figure, and movement of humans beyond this special meaning end up – via the thesaurus – in a collection area, where the dialectal lexis is grouped according to word fields. They can gauge which body images were idealized or disapproved of and how linguistic ideas were formed and used analog to fauna, flora, and artifacts. One of the documents (MWW-B011-009-083-001) contains an excerpt from the educational journal *Monatsschrift von und für Mecklenburg*, published in Schwerin from 1788 to 1791. Under the heading "Zoology," it is stated that the designations "Bollahrs" and "Bollehrschen" are the Kluthuhn (see Fig. 3).


1) aus der Zoologie.

Aar, (Entenſtößer, Fiſchahr) Falco haliaëtus. In der plattdeutſchen Bibel wird das Wort Adler ſtets durch Aarn überſezt. b)

Bollahrs, Bollehrſchen, (Kluthuhn, ein Huhn, das keinen Schwanz hat.) —

Borg, (ein verſchnittener Eber) Sus ſcroſa mas, caſtratus.

Brömm'r **), (Bull) Bos taurus.

**Fig. 3:** Excerpt from the magazine *Monatsschrift von und für Mecklenburg* (1791)

Once the dictionary project is completed, users can go from the above mentioned notes to the lemma "Bollmoors" and even more precisely to the sublemma "bollnoorst" of the digitized *Wossidlo-Teuchert* (vol. 1, col. 1003–1004). Here one finds the Latin name (*gallus ecaudatus*, see Fig. 4) and some linguistic form variants. Or users come from the dictionary via lemma entry and end up in the source documents presented by WossiDiA. Via the Trier dictionary network[6], more regional language dictionaries, Grimms' *Deutsches Wörterbuch*, Wander's proverb-dictionary, and other lexica connected to this portal can be searched with the help of a hyper lemma bar. The exchange processes are partly similar to those with the ISEBEL harvester.

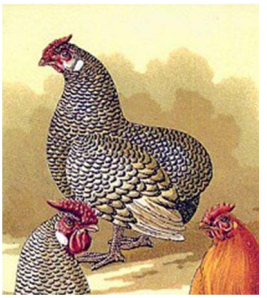


**Fig. 4:** *Gallus ecaudatus*, the "tailless" chicken. Detail from an illustration by Jean Bungartz (1891)

6 https://woerterbuchnetz.de (March 10, 2023).

Let us take a closer look at the local places made visible by the hypergraph extract (see Fig. 5). Wossidlo compiled the places of action of local legends in separate directory units. In the bundle about the area “Rambow” the tale of the witch who transforms herself into a tailless chicken can be found again (ZAW-A803-001-014). It is told in more detail and therefore represents the source document from which the above-mentioned fragmented version (ZAW-B703-004-005) was derived. In the detailed version, the informant states that the daughter herself (!) said that her mother had often turned into a tailless chicken and flew to Rotenmoor, where she had not left the people’s side, expressed by the phrase “he/se sitt up’t Heck” (*he/she is sitting on the hedge*) (see MWW-H036-004-030-016). The legend was recorded in June 1909.

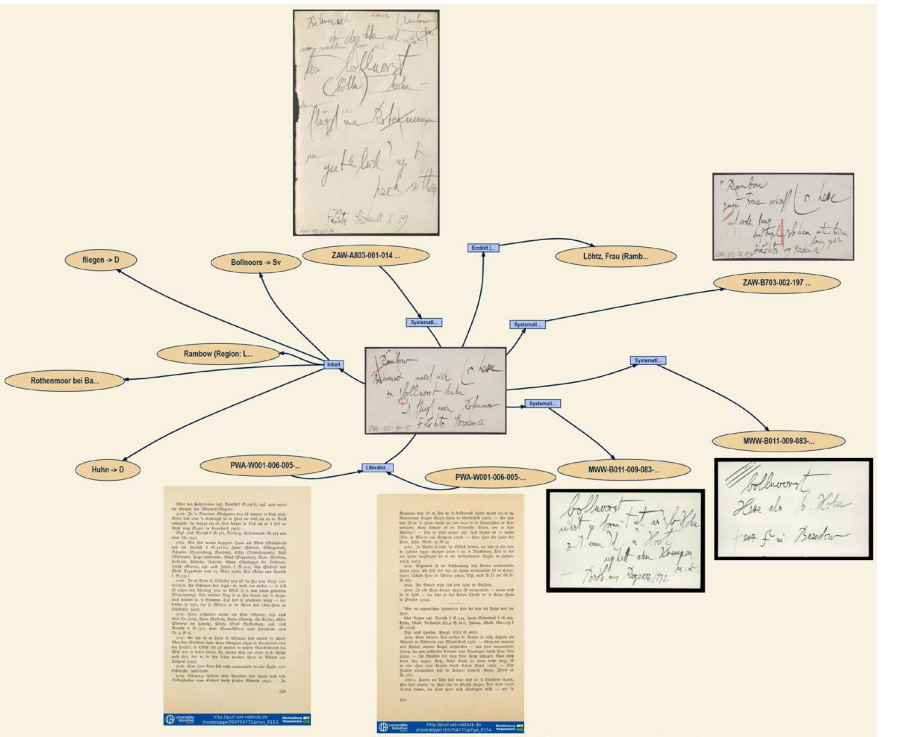



**Fig. 5:** Hypergraph extract

Scanned pages of published texts can be linked directly with their source references (see Fig. 5). Thus, one arrives via the slip of paper mentioned above (witch transforms herself into the tailless chicken, ZAW-B703-004-005) at Wossidlo’s *Meck-*

*lenburgische Sagen* (Wossidlo 1939, vol. 2, 329–330, no. 1029, see PWA-W001-006-005-033). However, only a brief reference to the motif variant can be found here. The female narrator describes her presumed birthplace Rambow as an "Hexenuurt" (ZAW-A803-001-012), i.e., as a "witch place" or "witch village". Stöckensee, the place of recording, is only about ten kilometers away. Where she got the variant of the tailless chicken remains in the dark.

Now to the interpretation of the enriched fieldnote (see Fig. 5). *Gallus ecaudatus* is a strange figure but, as a pet, familiar and not very scary. The initial version notes that the daughter refers to her mother as a witch; however, no damage spell is mentioned. Some questions might be evoked: Was the daughter annoyed that her mother, like a silly, curious chicken, often ran into the neighboring village and kept herself and the people from working? So is it an everyday conflict between mother and daughter, which has formed into a rumor interwoven with supernatural elements? Thus, the hypergraph extract may stimulate (not answer!) the question of how the reception of legends around 1900 changed from still believed to less believed stories. Users may then consider whether it would be promising to continue the hypergraph-based search.

The data discussed can also help explore the history of legend research and its paradigm around 1900. It is well known that the Brothers Grimm regarded legends as a kind of historical testimony which would go beyond historical sources. The followers of the *Mythologische Schule* (School of mythology), inspired by Jacob Grimm, exaggerated this widely speculative view. Wossidlo has also been influenced by this paradigm when interpreting legends. But he differs in that he recorded stories through field research. Therefore, it is interesting to take a closer look at how empirically collected data and its interpretation relate to each other.

While the female narrator saw Rambow as a witch place, Wossidlo counted it among the "old holy places" (Wossidlo 1919, 46) because the tiny village keeps one of the oldest churches of Mecklenburg. Therefore, this narrator and her demonological repertoire of tales attracted the collector's special attention for reconstructing mythological relics. It fits that Wossidlo marked the narrator's statement that Rambow is a witch place ("Hexenuurt," see ZAW-A803-001-012) by the abbreviation "ss" (= *sua sponte*). This means that the narrator told this of her initiative. This way, Wossidlo sought to avoid or mark suggestive questions concerning his research theses, which he was aware of in the field.

One of Wossidlo's legend research aims was to reconstruct the Germanic and Slavic legend traditions and their mixtures. Today we know that this is very difficult and often impossible. Many of the witch legends widespread in Mecklenburg seemed unsuitable to Wossidlo because they would only try to explain the damaging events of everyday life. However, in the case of "witch villages," he suspected that the pagan (Slavic) priesthood had once lived there (Wossidlo 1939, vol.

2: 307). This is highly speculative and corresponds to mythological interpretation processes. We do not intend to offer an answer but make it understandable how this interpretation came about. It is not entirely stupid, considering the struggle for pagan worship, central to public life, included suppression, and contempt of the conquered cultures. And that, in a way, fits in with the witch hunt. It should also be borne in mind that the Slavs migrated away from Mecklenburg in the late Middle Ages, and that part of the population has been assimilated. To trace Wossidlo's scholarly discourse on "witch villages," users can visit appropriately tagged excerpts from publications of the scientific community at that time (see ZAW-A116-007-002-001).

The former "Burgwall" (castle wall) research assumed that a Slavic sanctuary had previously been situated at Rambow. The Christian church's origin is believed to be in the fifteenth century. Due to the lack of historical evidence, Wossidlo took field names and prehistoric findings into account. Nearby Rothenmoor, where the witch flew, a former Slavic castle wall was found. Users could be informed about this data via an annotation module or a link to a cultural landscape. It should be borne in mind that this is not intended to revive the paradigm of legends' continuities.

With the help of WossiDiA, it may be worthwhile for archaeologists (who do research on castle walls for example) to follow the paths of research on former legends in the context of recorded field names. In WossiDiA 24,204 mostly georeferenced field names are currently entered in the hypergraph (cf. Meyer 2021), which can be widely related to local legends. For example, getting hints about soil cultivation or natural conditions may be interesting. This is especially true since such requests require little research time. Many other research scenarios are conceivable. For instance, the relationship between religious legends and customs (e.g., in tales about the dead and the death) could be enlightened. These data could be computatively compared with surveys (like those of Teuchert, Wossidlo, and the *Atlas der deutschen Volkskunde*, see Schmitt 2021: 131). Thematically appropriate surveys of today (e.g., about funeral rites) could be included (to stay with the example: to gain clues to changing images of death). Source criticism of the older surveys should be included in appropriate comment fields.

## 7 Conclusion

The WossiDiA project shows how a thematically diverse tradition archive can be transferred into a digital research environment. We have shown that modeling by hypergraphs is suitable for mapping the folklore collection's idiosyncratic struc-

ture. By WossiDiA the large amount of fieldnotes can be depicted and yet the documents can individually speak by networking them into larger units. It is no exaggeration to say that via WossiDiA and its hypergraph technique, some synapses of the collector's scientific train of thought can be enlightened. This will work even better after we have entered more metadata and transcriptions.

It is planned to facilitate the usage of WossiDiA with different templates, including data mining tools. Selected nodes and hyperedges can already be visualized. The applications must be taught to users more didactically and presented at least bilingually (including in English). The development of ISEBEL is partly a model for this. Our team often discusses what the digital archive must be able to do and what can be expected from users. Users should report their research results back to WossiDiA. For this, a particular workflow must be developed. Be it as it may: Even the mere public delivery of structured archival texts provides considerable added value. For narrative research, what was told and recorded in the field, retains its value anyway.

# 8 Literature

Becker, Siegfried: Das „Zentrale" am Archiv für Erzählforschung. In: Volkskundliche Großprojekte. Ihre Geschichte und Zukunft. ed. Christoph Schmitt. Münster/New York/München et al. 2005, 63–72.

Dietzsch, Ina: Zwischen Mathematik und Poesie. Praxen der Herstellung und Veröffentlichung volkskundlichen Wissens. In: Horizonte ethnografischen Wissens. Eine Bestandsaufnahme. eds. Ina Dietzsch/Wolfgang Kaschuba/Leonore Scholze-Irrlitz. Köln/Weimar/Wien 2009, 16–39.

Emerson, Robert M./Fretz, Rachel I./Shaw, Linda L.: Writing ethnographic fieldnotes. 2. Ed. Chicago/London 2011.

Enzyklopädie des Märchens. Bd. 15. Verzeichnisse, Register, Corrigenda. Berlin/Boston 2015.

Göttsch-Elten, Silke: Richard Wossidlo – ein Pionier der wissenschaftlichen Volkskunde. In: Kieler Blätter zur Volkskunde 41 (2009) 9–20.

Henßen, Gottfried: Mecklenburger erzählen. Märchen, Schwänke und Schnurren aus der Sammlung Richard Wossidlos. Berlin 1957.

Krohn, Kaarle: Erste Mitteilung des Folkloristischen Forscherbundes „FF" (23. November 1907). Helsinki 1907, 1–5.

Meyer, Holger/Schering, Alf-Christian/Schmitt, Christoph: WossiDiA – The Digital Wossidlo Archive. In: Corpora ethnographica online. Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet. Münster/New York/München et al. 2014, 61–84. https://doi.org/10.18453/rosdok_id00002265.

Meyer, Holger/Schering, Alf-Christian/Heuer, Andreas: The Hydra.PowerGraph System – Building Digital Archives with Directed and Typed Hypergraphs. In: Datenbank-Spektrum 17,2 (2017) 113–129.

Meyer, Holger/Mukbhil, Rasha/Schering, Alf-Christian: The Hydra.PowerGraph Data Definition, Manipulation, and Query Language GrafL. CS-01-17. The University of Rostock, CS Department 2017.

Meyer, Holger: Ein digitales Flurnamenarchiv für Mecklenburg-Vorpommern. In: Von Mund zu Ohr via Archiv in die Welt. Beiträge zum mündlichen, literarischen und medialen Erzählen. eds. Petra Himstedt-Vaid/Susanne Hose/Holger Meyer/Siegfried Neumann. Münster/New York 2021, 119–137.

Neumann, Siegfried: Volksschwänke aus Mecklenburg. Aus der Sammlung Richard Wossidlos. ed. Siegfried Neumann. Berlin 1963.

Neumann, Siegfried: Mecklenburgische Volksmärchen. ed. Siegfried Neumann. Berlin 1971.

Peuckert, Will-Erich/Lauffer, Otto: Volkskunde. Quellen und Forschungen seit 1930. Bern 1951.

Röhrich, Lutz: Deutscher Sagenkatalog X. Der Tod und die Toten (zusammen mit Ingeborg Müller). Einführung und Katalog. In: Deutsches Jahrbuch für Volkskunde 13 (1967) 346–397.

Ryan, Catherine/Mac Cárthaigh, Críostóir: Legacy folklore collections and the use of thesauri. In: Visions and Traditions. Knowledge Production and Tradition Archives (= FFC 315). eds. Lauri Harvilahti et al. Helsinki 2018, 297–322.

Schering, Alf-Christian: "WossiDiA History" & PowerGraph … "Make WossiDiA Efficient Again." In: Von Mund zu Ohr via Archiv in die Welt. Beiträge zum mündlichen, literarischen und medialen Erzählen. eds. Petra Himstedt-Vaid/Susanne Hose/Holger Meyer/Siegfried Neumann. Münster/New York 2021, 139–148.

Schmitt, Christoph: Zettelwerkstatt. Feldforschungsbasierte Wissenszirkulation um 1900 und die Praxis papierner Gelehrtenmaschinen am Fallbeispiel des „Volksforschers" Richard Wossidlo. In: Volkskunde in Sachsen 27 (2015) 7–47.

Schmitt, Christoph/Tangherlini, Timothy R.: Folklore Archives Online. Zur Sichtbarmachung, Auswertbarkeit und Interoperabilität einer dänischen und einer nordostdeutschen Sammlung. In: Jahrbuch für Europäische Ethnologie 13 (2018) 181–204.

Schmitt, Christoph: From Idiosyncratic Index-Card Machines to Digital Folklore Archives. The Case of WossiDiA. In: Folkloristics in the Digital Age (= FFC 316). eds. Pekka Hakamies/Anne Heimo. Helsinki 2019, 132–157.

Schmitt, Christoph: Data Sharing in der kulturwissenschaftlichen Erzählforschung. Vernetzungsszenarien digitalisierter Sammlungen mit digitalen Editionen. In: Analytische Fantasie. Von narrativen Welten zum guten Altern. Eine Festschrift für Harm-Peer Zimmermann. eds. Maximilian Jablonowski/Valerie Keller/Simone Stiefbold/Malte Völk. Weimar 2022, 128–137.

Schneidewind, Gisela: Herr und Knecht. Antifeudale Sagen aus Mecklenburg. Berlin 1960.

Thompson, Stith: Motif-Index of Folk-Literature. Vol. 1–6. Bloomington 1955–1958.

Uther, Hans-Jörg: The Types of International Folktales. A Classification and Bibliography, based on the System of Antti Aarne and Stith Thompson (= FFC 284–286). Helsinki 2004.

Wossidlo/Teuchert Mecklenburgisches Wörterbuch. Bd. 1–7. Berlin/Neumünster 1942–1992.

Wossidlo, Richard: Altheilige Stätten in Mecklenburg. In: Mecklenburg 14,1 (1919) 41–54.

Wossidlo, Richard: Mecklenburgische Sagen. Vol. 1–2. Rostock 1939.