Creation of Objects and Concordances for Knowledge Processing and Advanced Computing

Claus-Peter Rückemann
Westfälische Wilhelms-Universität Münster (WWU),
Leibniz Universität Hannover,
North-German Supercomputing Alliance (HLRN), Germany
Email: ruckema@uni-muenster.de

***Abstract*—This paper presents the summary of the main results from creating objects and concordances used for advanced processing and computing based on knowledge resources. Today big data collections and resources sadly combine one or more deficits of being unclassified, unstructured, isolated, and weakly developed on the one hand and in consequence only accessible with insufficient simplistic means on the other hand. The goal of this research is to create new classification features, structures, and components, which can be flexibly used with multi-disciplinary, multi-lingual long-term knowledge resources. The focus is to develop new extended facilities for universal long-term knowledge resources beyond the simple and isolated use of knowledge and data.**

Keywords–Creation of Objects and Concordances; Universal Decimal Classification; Sustainability; Knowledge Processing; Advanced Computing.

I. INTRODUCTION

This paper summarises the results from the development of advanced object features for long-term knowledge resources, which can be used for universal documentation and consequent purposes. For the consequent purposes like knowledge discovery, further creation and development of knowledge resources, visualisation, and education we can implement suitable, data-centred structures being usable with dynamical computing, knowledge processing, advanced computing.

Up to now the world of increasingly big data is limited to data and data collections, which are rapidly growing in quantity, mostly even growing in storage requirements instead of knowledge only. Besides the data being unstructured, isolated, and often inconsistent in content and form it is also missing quality and essential features for conceptional knowledge like classification. One consequence is that in the last decades the means for accessing and handling data have not changed a lot regarding the content, context, and a next generation of features and quality.

The focus of this research aims on the sustainable creation of long-term facilities for integrated documentation and applicability. The facilities provide advanced features for processing of knowledge as well as for flexible computing. Therefore, the creation and long-term care for suitable knowledge objects is a central issue. The knowledge resources have to be able to document any knowledge and data, e.g., factual and procedural knowledge, which require vertical as well as horizontal scalability for individual and subsequent use.

In order to cope with the deficits this architecture can integrate structured and unstructured data, support universal classification and concordances and it can enable advanced knowledge processing, like parallel processing and dynamical visualisation. This paper presents the up-to-date research results from the creation of long-term knowledge resources' components, structures, and workflows for advanced processing and computing – and in the end most important on the long run, fostering the investments in the development of the data itself. Therefore, the major contributions of this research are the content and context on the one hand and on the other hand the new practical insights on improving the state-of-the-art of long-term documentation and application of universal knowledge.

This paper is organised as follows. Section II introduces the state-of-the-art and the motivation for creating an overall system. Sections III and IV present the data-centric architecture, the implementation, and resources, especially the creation of objects and concordances. Section V shows the implemented features for knowledge processing and computing and Section VI provides an evaluation. Section VII presents the main results, summarises the lessons learned, conclusions and future work.

II. STATE-OF-THE-ART AND MOTIVATION

Many developments have contributed to the state-of-the-art on knowledge processing and discovery. Also, many developments have provided new technical means to cope with the new developments in computing architectures and services. In the context of knowledge processing, discovery, and search we are often faced with "prominent" examples like Internet search engines, library search engines, specialised expert systems, and maybe social media systems. If at all then there are some interfaces, e.g., for automated requests or web-service creation.

The algorithms applicable to this kind of 'art' are very limited and often not sufficient in delivering a reasonable quality for requested results or for following advanced goals. Therefore, various concepts, developments, and approaches have been created, addressing different aspects and special purposes. Nevertheless, these tools, classifications, and algorithms only try to handle the symptoms of the state of the data.

The approaches cover in-depth classifications and handling, e.g., library classification specialised on geological publications [1], handling historical geographic resources, especially in library context [2], and international patent classification [3].

 ISBN: 978-1-61208-416-9

The algorithms touch processing and automation, e.g., statistical models for online text classification [4] and automation with a classification [5]. The discussions and analyses range from research aspects to reliability and non-disciplinary approaches, e.g., classification as a research tool [6], reliability of diagnostic classification [7], and the Universal Decimal Classification (UDC) [8] as a non-disciplinary [9] universal [10] classification system, and legal and general aspects within Information Science, Security, and Computing [11].

In depth, aspects of mapping, organisation and multi-lingual data have been discussed, e.g., simple mapping between a classification and an "index" [12], simple conceptual methods for using classification in libraries [13], knowledge organisation [14], multi-lingual lexical linked data cloud [15]. In principle, any multi-disciplinary data resources may be used, e.g., projects like Europeana [16], European Cultural Heritage Online (ECHO) [17] or World Digital Library (WDL) [18]. Although these examples are focussed on providing special information they lack in sufficient content, organisation, and structure.

The main motivation for this research was the lack of multi-disciplinary data-centric approaches, which can be used for long-term creation of knowledge and scalable implementations. The data used here is based on the content and context from the knowledge resources, provided by the LX Foundation Scientific Resources [19]. The LX knowledge resources' structure and the classification references [20] based on UDC [21] are essential means for the processing workflows and evaluation of the knowledge objects and containers. Both provide strong multi-disciplinary and multi-lingual support. For this part of the research all small unsorted excerpts of the knowledge resources objects only refer to main UDC-based classes, which for this part of the publication are taken from the Multilingual Universal Decimal Classification Summary (UDCC Publication No. 088) [8] released by the UDC Consortium under the Creative Commons Attribution Share Alike 3.0 license [22] (first release 2009, subsequent update 2012). The analysis of different classifications, development of concepts for intermediate classifications, and experiences from case studies from the research conducted in the Knowledge in Motion (KiM) long-term project [23] have contributed to the application of UDC and different classifications and concordance schemes in the context of knowledge resources.

The following term definitions for object, container, and matrix can be helpful in this context. An object is and entity of knowledge data being part of knowledge resources. An object can contain any documentation, references, and other data. Objects can have an arbitrary number of sub-objects. A container is a collection of knowledge objects in a conjoint format. A matrix is a subset of the entirety, the "universe", of knowledge. A workflow can consist of many sub-workflows each of which can be based on an arbitrary number of knowledge matrices. The output of any sub-workflow or workflow can be seen as an intermediate or final result matrix.

III. Data-centric implementation

In order to concentrate on the challenges of the data itself so-called data-centric, data-defined, document-oriented or document-centric approaches have been developed. This went along with extending features like Structured Query Language (SQL) and "Not only SQL" (NoSQL) [24], e.g., via MySQL [25] and MongoDB [26] and in consequence [27] also in bridging relational and data- or document-centric approaches [28]. The very minimalistic "map" and "reduce" functions approach of MapReduce [29], which attracted many quick and simple solutions is a nice example for building simple workflow elements. As the knowledge resources' approach [19] is even much more general [20] it allows for arbitrary measures and also for processing implementing map and reduce functions, which can be based on the creation of objects and concordances.

Regarding a distributed computer system theoretical computer science can state the CAP (Consistency, Availability, Partition tolerance) theorem. In condensed form this means: *Consistency*: All nodes see the same data at the same time, *Availability*: A guarantee that every request receives a response about whether it succeeded or failed, *Partition tolerance*: The system continues to operate despite arbitrary message loss or failure of part of the system. Accordingly, learning from decades of case-studies, regarding the long-term knowledge and information sciences we can state a "CLU" theorem:

- *Consistency*: All knowledge in context used is neither in contradiction to other knowledge in context nor discording within its content,
- *Long-term sustainability*: Data-centric architectures, the core knowledge resources can be used for an arbitrary number of different implementations,
- *Universal documentation*: Documentation is supported for any knowledge and data, e.g., factual, conceptual, procedural, and metacognitive knowledge.

There is no direct reasonable equivalent for the P and A aspects. Besides the consistency, the items much more important are the long-term sustainability and universal documentation aspects. This includes the requirements for any type of knowledge as well as its multi-lingual documentation and features.

IV. Implementation and resources

The implementation for dynamical visualisation and computation is based on the framework for the architecture for documentation and development of advanced scientific computing and multi-disciplinary knowledge [30]. The architecture implemented for an economical long-term strategy is based on different development blocks. Figure 1 shows the three main columns: Application resources, knowledge resources, and originary resources. The central block in the "Collaboration house" framework architecture [20], are the knowledge resources, scientific resources, databases, containers, and documentation (e.g., LX [19], databases, containers, list resources). These can be based on and refer to the originary resources and sources (photos, scientific data, literature). The knowledge resources are used as a universal component for compute and storage workflows. Application resources and components (Active Source, Active Map, local applications) are implementations for analysing, utilising, and processing data and making the information and knowledge accessible. The related information, all data, and algorithm objects presented are copyright the author of this paper, LX Foundation Scientific Resources [19], all rights reserved. The structure and the classification references based on the LX resources and UDC, especially mentioning the well structured editions [8] and the multi-lingual features [21], are essential means for the processing workflows and evaluation of the knowledge objects and containers. Both provide strong multi-disciplinary and multi-lingual support.

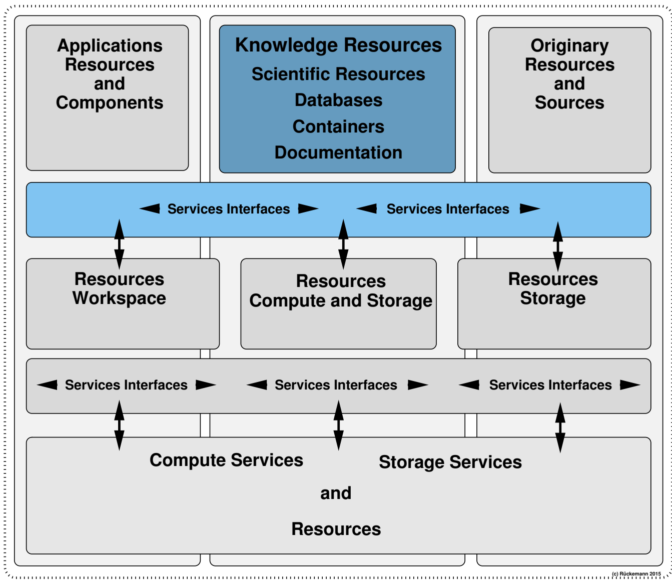

Figure 1. Architecture: Columns of practical dimensions. The knowledge resources are the central component within the long-term architecture.

The three blocks are supported by services' interfaces. The interfaces interact with the physical resources, in the local workspace, in the compute and storage resources the knowledge resources are situated, and in the storage resources for the originary resources.

All of these components do allow for advanced scientific computing and data processing, as well as the access of compute and storage resources via services interfaces. The resources' needs depend on the application scenarios to be implemented for user groups.

A. Creation of objects

Practical creation of objects has shown to be most efficient when three different categories of creation are considered:

- Manually created objects,
- Hybrid (semi-automatically) created objects, and
- Automatically created objects.

In any case creating objects is supported by universal classification, e.g., references to UDC. Therefore, that can also be applied for the creating concordances with objects. The listing in Figure 2 shows an instance of a simple object excerpt.

```
Vesuvius [Vulcanology, Geology, Archaeology]:
 (lat.) Mons Vesuvius.
 (ital.) Vesuvio.
 (deutsch.) Vesuv.
 Volcano, Gulf of Naples, Italy.
 Complex volcano (compound volcano).
 Stratovolcano, large cone (Gran Cono).
 Volcano Type: Somma volcano,
 VNUM: 0101-02=,
 Summit Elevation: 1281 m.
 The volcanic activity in the region is observed by the
 Oservatorio Vesuviano. The Vesuvius area has been
 declared a national park on 1995-06-05.
 The most known antique settlements at the Vesuvius are
 Pompeji and Herculaneum.
 Syn.: Vesaevus, Vesevus, Vesbius, Vesvius
 s. volcano, super volcano, compound volcano
 s. also Pompeji, Herculaneum, seismology
 compare La Soufrière, Mt. Scenery, Soufriere
 ...
```

Figure 2. Processed instance of a simple object (excerpt).

The listing in Figure 3 shows an instance of a simple container entry excerpt from a volcanological features container.

```
...
CONTAINER_OBJECT_EN_ITEM: Vesuvius
CONTAINER_OBJECT_DE_ITEM: Vesuv
CONTAINER_OBJECT_EN_PRINT: Vesuvius
CONTAINER_OBJECT_DE_PRINT: Vesuv
CONTAINER_OBJECT_EN_COUNTRY: Italy
CONTAINER_OBJECT_DE_COUNTRY: Italien
CONTAINER_OBJECT_EN_CONTINENT: Europe
CONTAINER_OBJECT_DE_CONTINENT: Europa
CONTAINER_OBJECT_XX_LATITUDE: 40.821N
CONTAINER_OBJECT_XX_LONGITUDE: 14.426E
CONTAINER_OBJECT_XX_HEIGHT_M: 1281
CONTAINER_OBJECT_EN_TYPE: Complexvolcano
CONTAINER_OBJECT_DE_TYPE: Komplex-Vulkan
CONTAINER_OBJECT_XX_VNUM: 0101-02=
```

Figure 3. Processed instance of a simple container entry (excerpt).

The excerpts have been processed with the appropriate `lx_object_volcanology` and `lx_container_volcanology` interfaces, selecting a number of items and for the container also items in English and German including a unique formatting. The resources' access and processing can be done in any programming language, assuming that the interfaces are implemented. For example, combining scripting, filtering, and parallel programming can provide flexible approaches.

B. Creation of concordances

Many disciplines and large fields of application have developed and used individual adapted frameworks of conceptual knowledge for their purposes. The reasons have been multifold, in that cases either developing a universal approach was too demanding or a distinction for certain reasons might have been considered adequate. In many cases, various classifications required to be "compared" and to be used together.

However, when developing content with conceptual knowledge and classifications sooner or later also the individual classifications get in the focus of development and may require to be "mapped". In most cases this can be done with the means of concordances, for example, concordances with classification in medicine and health [31] or the creation concordances between two classifications systems [32], in

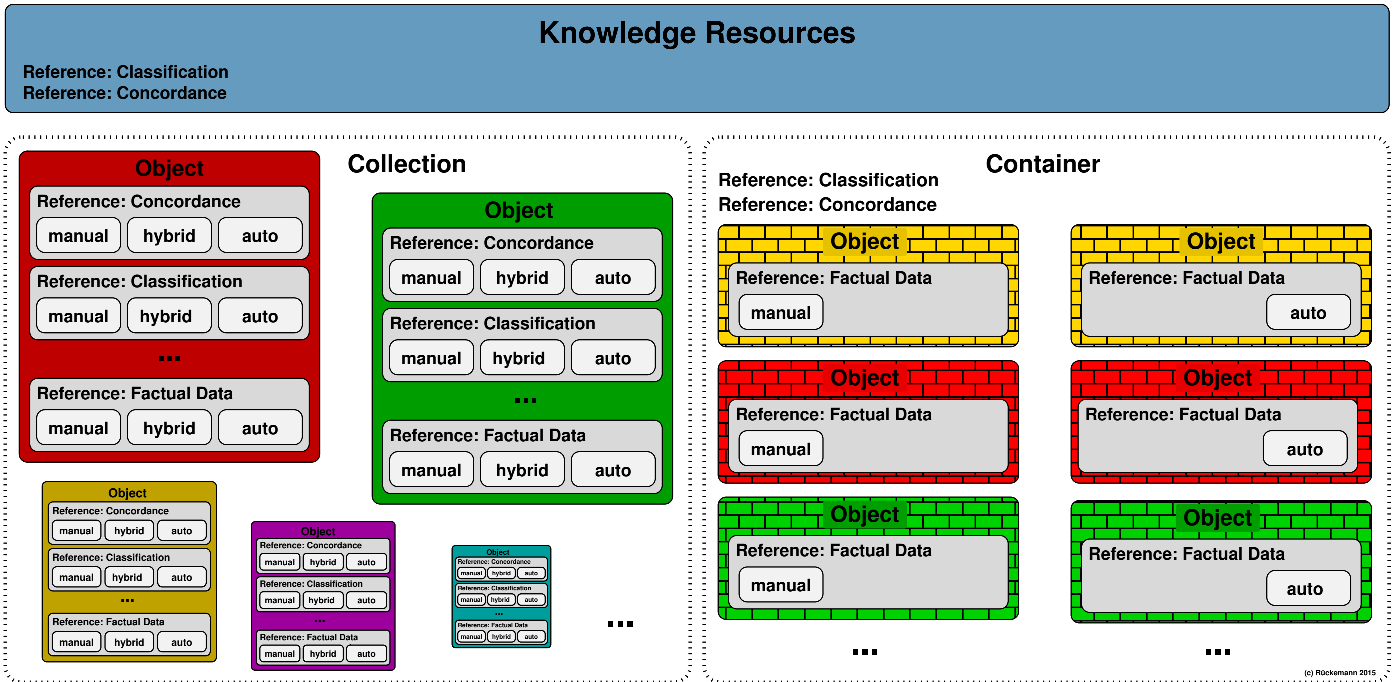

Figure 4. Resources and objects: Selected knowledge resources' objects containing references for concordances and classifications in collections and containers. In this case, the excerpt shows a distinct handling of manually, hybrid, and automatically created data, especially regarding classifications and concordances.

concordance projects like coli-conc [33] or for benefits in industry classification systems [34]. Therefore, the organisation of the resources and objects is significant for the long-term aspects and the vitality of the data. Taking advantage of the modular architecture of the overall resources (Figure 1) the main objectives are the knowledge resources, services, and interfaces, which are deployed for creating workflows. Figure 4 illustrates an excerpt of selected knowledge resources' objects. The selected objects are associated to collections and containers and contain references to concordances and classifications. The excerpt in this case shows a distinct handling of manually, hybrid, and automatically created data. The collection objects carry mostly only their individual conceptual knowledge, as there are concordances and classification, for example. The container objects are commonly similiar types of objects and structures where the container can carry a respective commonly valid conceptual knowledge for the container (symbolised in the figure by brick-structures for the objects). The respective knowledge resource, on the level integrating collections and resources, can also contain a respective commonly valid conceptual knowledge for the resource.

There are different ways of handling the processes for semi-automatically and automatically created concordances. With the main focus on processing and advanced computing we concentrate of the object and references side of the resources. This concept has shown vital benefits, which enables implementations with comparably high flexibility. Disciplines, services, and resources can be integrated in a very scalable way. Practical creation of concordances has shown to be most efficient when three different categories of creation are considered:

- Manually created concordances,
- Hybrid (semi-automatically) created concordances, and
- Automatically created concordances.

Manually-created-concordances is a type of concordances, which has resulted from manually inserting references from objects into a concordance instance. Hybrid-created-concordances is a type of concordances, which has resulted from a combination of manual and automated (semi-automatically) processes.

The processes may work on primary concordances or on any level of secondary data in order to support the creation of concordances. Automatically-created-concordances is a type of concordances, which have been generated, e.g., by an automated workflow process. This is mostly done for big data, which are used as quantity data and not due to their quality. In any way, an integration with the knowledge resources' references and structures is the target.

The workflows can contain several functions comparable to the map and reduce concept. A map function finds the data according to the criteria and creates a map result matrix. A reduce function does the appropriate operation on the map result matrix output. The listing in Figure 5 shows a simple object instance classification and concordances excerpt (Figure 2) from a volcanological object in a collection. The excerpt shows classification concordances in several different classifications as used in different disciplines. Possibly multiple views from different disciplines or author groups on a certain object are not shown in this reduced view but they can also hold the full spectrum of classifications and concordances.

```
...
UCC:UDC2012:551.21
UCC:UDC2012:551
UCC:UDC2012:902/908
UCC:MSC2010:86,86A17,86A60
UCC:LCC:QE521-545
UCC:LCC:QE1-996.5
UCC:LCC:QC801-809
UCC:LCC:CC1-960,CB3-482
UCC:PACS2010:91.40.-k
UCC:PACS2010:91.65.-n,91.
```

Figure 5. Classification and concordances excerpt of a simple object instance (knowledge resources collection).

The listing (Figure 6) excerpts classification and concordances of a (volcanological features) container (Figure 3).

```
UCC:UDC2012:551.21
UCC:UDC2012:551
UCC:UDC2012:551.2,551.23,551.24,551.26
UCC:UDC2012:902/908
UCC:MSC2010:86,86A17,86A60
UCC:LCC:QE521-545
UCC:LCC:QE1-996.5
UCC:LCC:QC801-809
UCC:LCC:CC1-960,CB3-482
UCC:PACS2010:91.40.-k
UCC:PACS2010:91.65.-n,91.
UCC:PACS2010:91.40.Ge,91.40.St,91.40.Rs,*91.45.C-,*91.45.
 D-,90
...
```

Figure 6. Classification and concordances excerpt of a simple container instance (knowledge resources container).

The differences in classification and concordances are resulting from the different level of detail in the collections and containers as well as in different potential of the various classification schemes to describe certain knowledge as can be seen from the different depth of classification. In integration, together the concordances can create valuable references in depth and width to complementary classification schemes and knowledge classified with different classification.

The term concordance is not only used in the simple traditional meaning. Instead, the organisation is that of a meta-concordances concept. That results from the use of universal meta-classification, which in turn is used to classify and integrate classifications [35]. The samples include simple classifications from UDC, Mathematics Subject Classification (MSC) [36], Library of Congress Classification (LCC) [37], and Physics and Astronomy Classification Scheme (PACS) [38]. For PACS the asterisk ($\star$) indicates entries from the “Acoustics Appendix / Geophysics Appendix”.

The Universal Classified Classification (UCC) entries contain several classifications. The UCC blocks provide concordances across the classification schemes. The object classification is associated with the items associated with the object whereas the container classification is associated with the container, which means it refers to all objects in the containers.

V. KNOWLEDGE PROCESSING AND COMPUTING

The advanced processing of knowledge resources benefits from a significant number of unique attributes in its elements. These attributes can be references, classification, keywords, textual content, links, and many more. The elements can consist of objects or collections of objects, the containers, integrating factual data with object information and structure.

Workflows for creating arbitrary result matrices (Figure 7) have been based on the organisation and object features (Figure 4) in the knowledge resources.

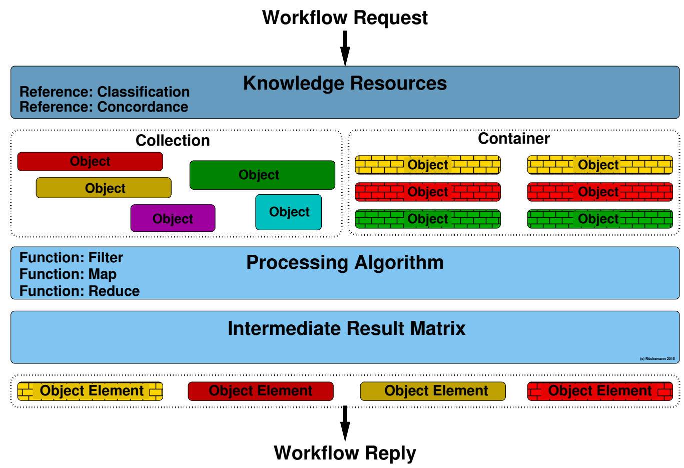

Figure 7. Creation of intermediate result matrices from resources and references (collections and containers) in reply to workflow requests.

The illustration shows that object information is gathered from the objects and references in collections and containers. Configurable algorithms like filters and mapping are then used in order to compute a result matrix. Here, the result matrix is considered “intermediate” because any of such workflows can be used in combination with other workflows, workflow chains or further processing.

For example, there is no “archaeology” in PACS, the concordances refer to resources including “archaeology” via some of the other schemes. MSC also does not contain a classification neither for volcanology or geology nor for associated features. Instead even the geophysics section classifying geological problems refers to computational methods. The above examples (Figures 5 and 6) also illustrate this. The concordances’ blocks allow to bridge between classifications and data resources, which can efficiently increase the available data pool size. Common options are in-depth computation with the container, or in-width with the general object collections. The concordances’ blocks allow to follow in-depth or in-width references within data resources, which efficiently supports to improve the quality of result matrices and the quantity of elements, which also impacts on scalability and efficiency of workflows. Table I shows the shares of items regarding processing and computing with the main steps at knowledge resources, processing algorithms, and intermediate result matrices for the Vesuvius/volcanology case (Figures 2–6).

 ISBN: 978-1-61208-416-9

TABLE I. PROCESSING AND COMPUTING WITHOUT (/w) AND WITH (w/) CLASSIFICATION & CONCORDANCES (VESUVIUS/VOLCANOLOGY CASE).

Items of Processing and Computing	*Values*	
	/w	*w/*
Knowledge resources		
Collection	10,000	200
Container	300	5
Processing algorithm		
Mean	750	230
String comparison	90,000	8,000
Associations	344	127
Phonetics	34	22
Weighting	296	84
Intermediate result matrix		
4 result matrix elements	120	20

The number of operations is based on subset of 100,000 collection and container objects from the knowledge resources, which have been accessed for the study. The number of items to be handled by the processing and computing for creating a comparable or higher quality result matrix have been much smaller in the major number of practical workflows when classification and concordances are included in the workflows. Especially, the primary number of requests on the collections and containers can be reduced. Consequently, the number of algorithm calls is reduced. The number of string comparisons and associated algorithms is most prominent here as the majority of objects in the resources contain text. Figure 8 shows an elementary sample workflow batch implementation of a generated caller script used for the processing parallelisation for the computing tasks, e.g., calling from Integrated System components like actmap [39].

```
#!/bin/bash
#PBS -A ruckema
#PBS -N PARA_Discover
#PBS -j oe
#PBS -l feature=mpp1
#PBS -l nodes=16:ppn=6
#PBS -l walltime=00:60:00
cd $PBS_O_WORKDIR
msub para_discover.sh
```

Figure 8. Generated workflow parallelisation with `PARA_Discover`.

Every instance of this sample Portable Batch System (PBS) script uses 16 compute nodes and 6 processors per node in order to execute a `para_discover` call for maximum 60 minutes walltime. A regular run with the above values requires about 5 minutes walltime per instance without and about 25 seconds with classification and concordances. With four times the nodes and cores we can handle about four times the subset data.

However, it is important to choose a right knowledge representation for universal long-term data. The Resource Description Framework (RDF) [40] is a simple example for representing Web data. In many cases, simple directed labeled graphs are not sufficient to represent knowledge. References to directed labeled or other kinds of representations should be possible.

The structure should provide an intuitive and flexible access to the data. There should be features for integrating any kind of external data, e.g., objects, references, links, from structured to unstructured data with the available data. The elementary means of accessing the data should be independent from a certain implementation or certain purpose. The integration, interfaces, and interchange of data should be provided in most sustainable ways. This means any kind of structure and references and conceptual knowledge representation can be integrated. For example, in case of Web data even RDF can be deployed for Uniform Resource Identifiers (URI) naming relationships between data at the "ends" of a link, which in simple context enables to use graph analytics even on powerful High End Computing resources.

VI. EVALUATION

As shown, objects and containers can carry complementary information and knowledge. The classifications and concordances feature a fuzzy bridging between resources, which allows modular in-depth as well as in-width workflows. In addition to that, workflows can require strongly adaptive code and algorithms. This may result in significant variations of runtime behaviour and resources' requirements. The workflows can integrate any objects for the processing, e.g., from collections and containers. These objects and their content may result from manual to automated origin. For example, the spectrum of creation includes use of classification, keywords, text analysis, and context analysis for the purpose of integration.

All the elements like classification, concordances, and factual data can result from manual, hybrid, and automatic processes. For example, Big Data resources can be automatically outfitted with classifications and concordances following the container components. The level of details in content, context, and structure is arbitrary and can be scaled defined by the focus of the creator of the respective data. Therefore, associated conditions can be used in workflows for weighting the types of processes and qualities involved.

In practice, during the processing and computing, the numbers of algorithm calls for requests on the collections and containers can be significantly reduced with considering classification and concordances in workflows even when creating result matrices of comparable or higher quality. There will always be non-automated resources, which might be the knowledge intensive ones. The knowledge review can also be supported by distributed authorities as well as by means of automation.

VII. CONCLUSION

The types of objects and concordances shown in this paper have been successfully created and further developed within the knowledge resources. These results have also been integrated into the knowledge resources. The workflows for creating the structures and the features for the advanced processing and computing based on these resources have been successfully implemented with in the last years. From this research, we have learned some major results.

Experiences with the creation and development of objects within the knowledge resources have resulted in the fact that the data-centric approach neither conflicts with the long-term aspects nor with the deployment of advanced processing and computing features. This way, it should be possible to keep knowledge persistent even under changes of technology and paradigms.

The integration of objects, classification, and concordances has provided new means of documenting and accessing knowledge as well as for the efficient application of computational means. The structure of the long-term multi-disciplinary and multi-lingual knowledge resources' components enables to easily integrate objects from collections and containers. In more depth the conceptual knowledge, e.g., the classification can improve the quality of the result matrices. It enables to integrate more objects via strong means of knowledge instead of statistics or pattern matching algorithms only.

The implementation of the concordances and workflows has shown that the integrability of objects regarding multi-disciplinary and multi-lingual aspects has improved. The introduction of a universal classification and concordances is an excellent means of breaking up the isolated state of knowledge resources' content and associated data. In this context, creating concordances mainly contribute in two ways. On the one hand, concordances enable to considers different views of different and even special disciplines with the knowledge processing. On the other hand, concordances can be used to build bridges between isolated data resources.

The flexibility of the knowledge processing benefits from the advanced organisation of the data, which enables various scalable computational means for implementing directed graphs to fuzzy links, for which High End Computing resources can be deployed. Future work will be focussed on intelligent ways how classifications, concordances, and references, e.g., intermediate classifications, can be created and deployed for the use with long-term knowledge resources, classification, and workflows.

ACKNOWLEDGEMENTS

We are grateful to all national and international partners in the GEXI cooperations for their support and contributions. Special thanks go to the scientific colleagues at the Gottfried Wilhelm Leibniz Bibliothek (GWLB) Hannover, especially to Dr. Friedrich Hülsmann, for prolific discussion, inspiration, and practical case studies. Many thanks go to the scientific colleagues at the Leibniz Universität Hannover, especially to Mrs. Birgit Gersbeck-Schierholz for practical multi-disciplinary case studies and discussions. We are grateful to the Knowledge in Motion (KiM) long-term project and its members, DIMF, for the contributions on the development and application of classifications.

REFERENCES

[1] R. S. Sasscer, U.S. Geological Survey library classification system. U.S. G.P.O., 1992, USGS Bulletin: 2010.

[2] E. Dodsworth and L. W. Laliberte, Eds., Discovering and using historical geographic resources on the Web: A practical guide for Librarians. Lanham: Rowman and Littlefield, 2014, ISBN: 0-8108-914-1.

[3] "Die Internationale Patentklassifikation (International Patent Classification, IPC)," 2014, Deutsches Patent- und Markenamt (DPMA), Germany, URL: http://dpma.de/service/klassifikationen/ipc/ [accessed: 2015-02-01].

[4] P. Cerchiello and P. Giudici, "Non parametric statistical models for on-line text classification," Advances in Data Analysis and Classification – Theory, Methods, and Applications in Data Science, vol. 6, no. 4, 2012, pp. 277–288, special issue on "Data analysis and classification in marketing" ISSN: 1862-5347.

[5] I. Dahlberg and M. R. Schader, Eds., Automatisierung in der Klassifikation, Proceedings, 7. Jahrestagung der Gesellschaft für Klassifikation e.V. (Teil 1), Königswinter/Rhein, Deutschland, 5.–8. April 1983. Indeks Verlag, Frankfurt a. M., 1983, ISBN: 3-88672-012-X, URL: https://openlibrary.org/books/OL21106918M/Automatisierung_in_der_Klassifikation [accessed: 2015-02-01].

[6] W. Gaul and M. Schader, Eds., Classification As a Tool of Research. North-Holland, Amsterdam, 1986, Proceedings, Annual Meeting of the Classification Society, (Proceedings der Fachtagung der Gesellschaft für Klassifikation), ISBN-13: 978-0444879806, ISBN-10: 0-444-87980-3, Hardcover, XIII, 502 p., May 1, 1986.

[7] J. Templin and L. Bradshaw, "Measuring the Reliability of Diagnostic Classification Model Examinee Estimates," Journal of Classification, vol. 30, no. 2, 2013, pp. 251–275, Heiser, W. J. (ed.), ISSN: 0176-4268 (print), ISSN: 1432-1343 (electronic), URL: http://dx.doi.org/10.1007/s00357-013-9129-4 [accessed: 2015-02-01].

[8] "Multilingual Universal Decimal Classification Summary," 2012, UDC Consortium, 2012, Web resource, v. 1.1. The Hague: UDC Consortium (UDCC Publication No. 088), URL: http://www.udcc.org/udcsummary/php/index.php [accessed: 2015-02-01].

[9] P. Cousson, "UDC as a non-disciplinary classification system for a high-school library," in Proc. UDC Seminar 2009, Classification at a Crossroads: Multiple Directions to Usability, 1992, pp. 243–252, URL: http://www.academia.edu/1022257/UDC_as_a_non-disciplinary_classification_system_for_a_high-school_library [acc.: 2015-02-01].

[10] A. Adewale, "Universal Decimal Classification (UDC): A Most For All Libraries," library 2.0, the future of libraries in the digital age, 2014, URL: http://www.library20.com/forum/topics/universal-decimal-classification-udc-a-most-for-all-libraries [accessed: 2015-02-01].

[11] "EULISP Lecture Notes, European Legal Informatics Study Programme, Institute for Legal Informatics (Institut für Rechtsinformatik, IRI), Leibniz Universität Hannover," 2014, URL: http://www.eulisp.de [accessed: 2015-02-01].

[12] F. Heel, "Abbildungen zwischen der Dewey-Dezimalklassifikation (DDC), der Regensburger Verbundklassifikation (RVK) und der Schlagwortnormdatei (SWD) für die Recherche in heterogen erschlossenen Datenbeständen – Möglichkeiten und Problembereiche," Bachelorarbeit im Studiengang Bibliotheks- und Informationsmanagement, Fak. Information und Kommunikation, Hochschule der Medien Stuttgart, 2007.

[13] T. Riplinger, "Die Bedeutung der Methode Eppelsheimer für Theorie und Praxis der bibliothekarischen und der dokumentarischen Sacherschließung," BIBLIOTHEK Forschung und Praxis, vol. 28, no. 2, 2012, pp. 252–262, 01/2004, DOI: 10.1515/BFUP.2004.252.

[14] S. A. Keller and R. Schneider, Eds., Wissensorganisation und -repräsentation mit digitalen Technologien. Walter de Gruyter GmbH, 2014, Bibliotheks- und Informationspraxis, ISSN: 0179-0986, Band 55, ISBN: 3-11-031270-0.

[15] E. W. De Luca and I. Dahlberg, "Die Multilingual Lexical Linked Data Cloud: Eine mögliche Zugangsoptimierung?" Information Wissenschaft & Praxis, vol. 65, no. 4–5, 2014, pp. 279–287, Deutsche Ges. f. Information und Wissen e.V. (DGI), Ed., De Gruyter Saur, ISSN: 1434-4653, (title in English: The Multilingual Lexical Linked Data Cloud: A possible semantic-based access to the Web?).

[16] "Europeana," 2015, URL: http://www.europeana.eu/ [accessed: 2015-02-01].

[17] Max Planck Institute for the History of Science, Max-Planck Institut für Wissenschaftsgeschichte, "European Cultural Heritage Online (ECHO)," 2015, Berlin, URL: http://echo.mpiwg-berlin.mpg.de/ [accessed: 2015-02-01].

[18] “WDL, World Digital Library,” 2015, URL: http://www.wdl.org [accessed: 2015-02-01].

[19] “LX-Project,” 2015, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX (Information) [accessed: 2015-02-01].

[20] C.-P. Rückemann, “Enabling Dynamical Use of Integrated Systems and Scientific Supercomputing Resources for Archaeological Information Systems,” in Proc. INFOCOMP 2012, Oct. 21–26, 2012, Venice, Italy, 2012, pp. 36–41, ISBN: 978-1-61208-226-4.

[21] “UDC Online,” 2015, URL: http://www.udc-hub.com/ [accessed: 2015-02-01].

[22] “Creative Commons Attribution Share Alike 3.0 license,” 2012, URL: http://creativecommons.org/licenses/by-sa/3.0/ [accessed: 2014-01-12].

[23] F. Hülsmann and C.-P. Rückemann, “Summary on Algorithms and Workflows,” KiMrise, Knowledge in Motion Winter Meeting, December 12, 2014, Knowledge in Motion, Hannover, Germany, 2014.

[24] S. Tiwari, Professional NoSQL. Wrox, John Wiley & Sons, Inc., 2011, ISBN: 978-0-470-94224-6.

[25] “MySQL,” 2015, URL: http://www.mysql.com/ [accessed: 2015-02-01].

[26] “MongoDB,” 2015, URL: http://www.mongodb.org/ [accessed: 2015-02-01].

[27] K. Chodorow, MongoDB: The Definitive Guide. O’Reilly Media, 2013, ISBN: 9781449344689.

[28] J. Roijackers and G. H. L. Fletcher, “On Bridging Relational and Document-Centric Data Stores,” in Proc. 29th British National Conf. on Databases (BNCOD 2013), Oxford, UK, July 8–10, 2013, Big Data, LNCS, vol. 7968, 2013, pp. 135–148, ISBN: 978-3-642-39466-9.

[29] J. Dean and S. Ghemawat, “MapReduce: Simplified data processing on large clusters,” in Proceedings of the 6th Conference on Operating Systems Design and Implementation (OSDI 2004), 2004, DOI: 10.1.1.163.5292.

[30] C.-P. Rückemann, “High End Computing Using Advanced Archaeology and Geoscience Objects,” International Journal On Advances in Intelligent Systems, vol. 6, no. 3&4, 2013, ISSN: 1942-2679, LCCN: 2008212456, URL: http://www.iariajournals.org/intelligent_systems/intsys_v6_n34_2013_paged.pdf [accessed: 2015-02-01].

[31] U. Balakrishnan, “Eine DDC-RVK-Konkordanz - Erste Erkenntnisse aus dem Gebiet Medizin & Gesundheit,” 2012, Projekt Colibri / DDC Teilprojekt coli-conc, URL: http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-82838 [accessed: 2015-02-01].

[32] I. Rauner, “Erstellung einer Konkordanz zwischen BK (Basisklassifikation) und RVK (Regensburger Verbundklassifikation) für das Fachgebiet Germanistik,” Master’s thesis, University of Vienna, Universitätslehrgang Library and Information Studies, 2010.

[33] U. Balakrishnan, “Das Konkordanzprojekt coli-conc,” 2012, Verbundzentrale des GBV, 13. November 2013, Göttingen, Germany, URL: https://www.gbv.de/cls-download/fag-erschliessung-und-informationsvermittlung/arbeitsdokumente-fag-ei/praesentation-zu-konkordanzen/at_download/file [accessed: 2015-02-01].

[34] “North American Industry Classification System (NAICS), Concordances,” 2014, URL: https://www.census.gov/eos/www/naics/concordances/concordances.html [accessed: 2015-02-01].

[35] F. Hülsmann and C.-P. Rückemann, “Classifying Classifications: Metaclassification as a New Method to Deal with Knowledge,” in Proceedings of The Fifth International Conference on Advanced Communications and Computation (INFOCOMP 2015), June 21–26, 2015, Brussels, Belgium. XPS Press, 2015, ISSN: 2308-3484, ISBN-13: 978-1-61208-416-9, (to appear).

[36] “Mathematics Subject Classification (MSC2010),” 2010, URL: http://msc2010.org [accessed: 2015-02-01].

[37] Fundamentals of Library of Congress Classification, Developed by the ALCTS/CCS-PCC Task Force on Library of Congress Classification Training, 2007, Robare, L., Arakawa, S., Frank, P., and Trumble, B. (eds.), ISBN: 0-8444-1186-8 (Instructor Manual), ISBN: 0-8444-1191-4 (Trainee Manual), URL: http://www.loc.gov/catworkshop/courses/fundamentalslcc/pdf/classify-trnee-manual.pdf [accessed: 2015-02-01].

[38] “Physics and Astronomy Classification Scheme, PACS 2010 Regular Edition,” 2010, American Institute of Physics (AIP), URL: http://www.aip.org/pacs [accessed: 2015-02-01].

[39] C.-P. Rückemann, “Active Map Software,” 2001, 2005, 2012, URL: http://wwwmath.uni-muenster.de/cs/u/ruckema (information, data, abstract) [accessed: 2012-01-01], URL: http://www.unics.uni-hannover.de/cpr/x/rprojs/en/index.html#actmap [accessed: 2015-01-02].

[40] W3 Consortium (W3C), “Resource Description Framework (RDF),” 2015, W3C Semantic Web, URL: http://www.w3.org/RDF/ [accessed: 2015-01-24].