Integration of Knowledge Resources and Advanced Association Processing for Geosciences and Archaeology

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 main research results on a major task for knowledge discovery: The integration of knowledge resources and advanced association processing. The research includes the integration of references to many types of objects and elements, e.g., adding conceptual knowledge via references to a universal classification. Creating sustainable multi-disciplinary knowledge resources and enabling advanced features for processing of associations is one of the major goals of long-term knowledge development and discovery, especially allowing the use of advanced association processing and computation facilities. Research has shown that discovery in geosciences and archaeology widely benefits from a multi-disciplinary approach. The paper also presents respective association processing results exploiting the integration of such geoscientific and archaeological knowledge resources components. The practical application scenario is based on content from a natural sciences and archaeology research and studies campaign at the ancient city of Kameiros, Greece. The created resources are providing content, structures, and features for exploiting computation facilities, especially a multitude of reference types. The focus is to integrate multi-disciplinary knowledge resources and association processing, which allows to extend and exploit long-term content discovery on a data-centric base and gain new insights.**

***Keywords*–Knowledge Discovery; Association Processing; Scientific Knowledge Resources; Universal Decimal Classification; Advanced Computing.**

I. INTRODUCTION

This extended research presents the main results on a major task for knowledge discovery: The integration of knowledge resources and advanced association processing. The research is based on the results from multi-disciplinary projects for the creation of objects and concordances, which are intended to be used for long-term knowledge creation knowledge processing, and advanced computing. The basic fundaments of the results were presented at the DigitalWorld / GEOProcessing 2016 conferences in Venice, Italy [1]. Previous research was conducted for creating knowledge resources and developing application components for supporting and providing advanced integrated systems for geoscientific, multi-disciplinary, and multi-lingual application scenarios. Existing data collections, unstructured and structured, combine a number of insufficient features and drawbacks, missing long-term aspects, support for multi-disciplinary conceptual knowledge, for classification, and for advanced and fuzzy methods like associations.

The purpose of the integration of the developed resources and components is to provide advanced knowledge object features, especially association processing features and computation in context with long-term multi-disciplinary and multi-lingual knowledge documentation and discovery. The new resources and application developments presented here are based on selected frameworks and resources, which have been created over the last two decades. The knowledge resources and Collaboration house framework [2] allowed for the implementation of multi-disciplinary, long-term knowledge resources and application components, for dynamical use as well as for complex and high end computation. The resulting components are used for universal and consistent documentation of knowledge and scientific research, and for consequent long-term purposes. These components are created using a universal classification [3], a flexible and portable all-purpose programming environment [4], and appropriate international standards [5]. In this case, for advanced association processing, new workflows had to be created and dynamically integrated into the framework components. Such implementation is possible if on the one hand the components' workflows allow a flexible integration of workflows, e.g., via scripting and compiled sources and on the other hand that structured knowledge resources can be extended for allowing a multitude of references types. The combination allows to create associations by making use of the available structures, processing, and computation facilities. For these purposes, the object and media knowledge resources and the framework components were basically extended to support a data-centric approach.

This data-centric approach includes the integration of references to many types of multi-disciplinary knowledge resources' objects and elements, e.g., adding conceptual knowledge via references to a universal classification. Therefore, the implementations and case study focus on spanning disciplines. In this case the major starting points are geosciences and archaeology and their referenced disciplines' context.

This paper is organised as follows. Section II presents the state of integration and frameworks with knowledge resources. Sections III and IV introduce the integration of workflows and reference types. Sections V and VI discuss the creation and processing of associations and how computational facilities can be exploited. Section VII presents and discusses an integration case study and implementation in context of geosciences and archaeology. Sections VIII and IX give an evaluation, present the main results and summarise the lessons learned, conclusions and future work.

II. State of integration and frameworks

The resources and implementations are based on the integration of three major components: An architecture framework, long-term, multi-disciplinary knowledge resources, and a mostly widely used universal classification framework. The architecture implemented for an economical long-term strategy is data-centric and based on 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 [6], is represented by the knowledge resources, scientific resources, object collections, containers, databases, and documentation (e.g., LX [7], collections, containers). These resources provide multi-disciplinary content, context, and references, including structured and unstructured data, factual and conceptual knowledge.

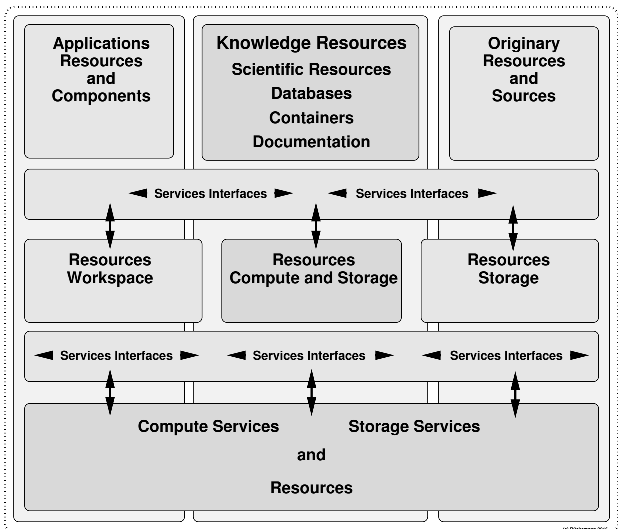

Figure 1. Architecture: The knowledge resources are the central component within the long-term architecture. Three major layers labelled (1), (2), (3).

The resources also refer to originary resources and sources (e.g., textual data, media data, photos, scientific data, literature). The knowledge resources are used as a universal component for compute and storage workflows. This feature can also be applied for supporting dynamical and ontology-based multi-agent, e.g., for production management as with the implementation supported by the European Framework Programme 7 (FP7) [8]. 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 [7], all rights reserved. The structure and the classification references based on the LX resources and UDC [9], especially mentioning the well structured editions [3] and the multi-lingual features [10], are essential means for the processing workflows and evaluation of the knowledge objects and containers. Both provide strong multi-disciplinary and multi-lingual support. 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 where the knowledge resources are situated, and in the storage resources for the originary resources. The layers or 'levels' are labelled (1), (2), and (3) within the architecture. (1) is associated with the disciplines creating and using knowledge resources, application resources, and originary resources, 'realia'. (2) is associated with the tasks and contributions of services providers. (3) is associated with the computer and storage resources provided by resources providers. 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. The framework allows to create any collaboration required for the development and operation of knowledge resources, required services, and High End Computing resources like compute and storage.

III. Integration of workflows

The integration of association processing workflows with the workflows for creating arbitrary result matrices is most flexible and efficient and was based on the organisation and object features (Figure 2) in the knowledge resources. Object details and definitions have been discussed with computational views [11]. Workflow steps are labelled with lowercase letters. Layers are labelled with numbers, primary layers are underlined. 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 to compute a result matrix. The result matrix is considered "intermediate" because any of such workflows can be used in combination with other workflows, workflow chains or further processing.

(a) Geoscientific Association Processing Workflow Request: A request for geoscientific knowledge resources is initiated from within a discovery workflow. Such request is created in level (2) within the architecture.
(b) Geoscientific Knowledge Resources: The respective resources are initialised for the workflow. The knowledge resources are located in level (1).
(c) Collections and containers: The collections and containers within the resources are provided.
(d) Association Processing Algorithms and Definitions: The algorithms and definitions for the association processing are called. The processing involves (1), (2), and (3), especially the last two.
(e) Association Processing Intermediate Result Matrix: An intermediate result matrix is created by the algorithms and definitions. The matrix creation involves (1), (2), and (3), especially (2).
(f) Geoscientific Association Processing Workflow Reply: Such reply is created in level (2) within the architecture.

Figure 3 illustrates selected knowledge resources' objects, focussing on references in collections and containers.

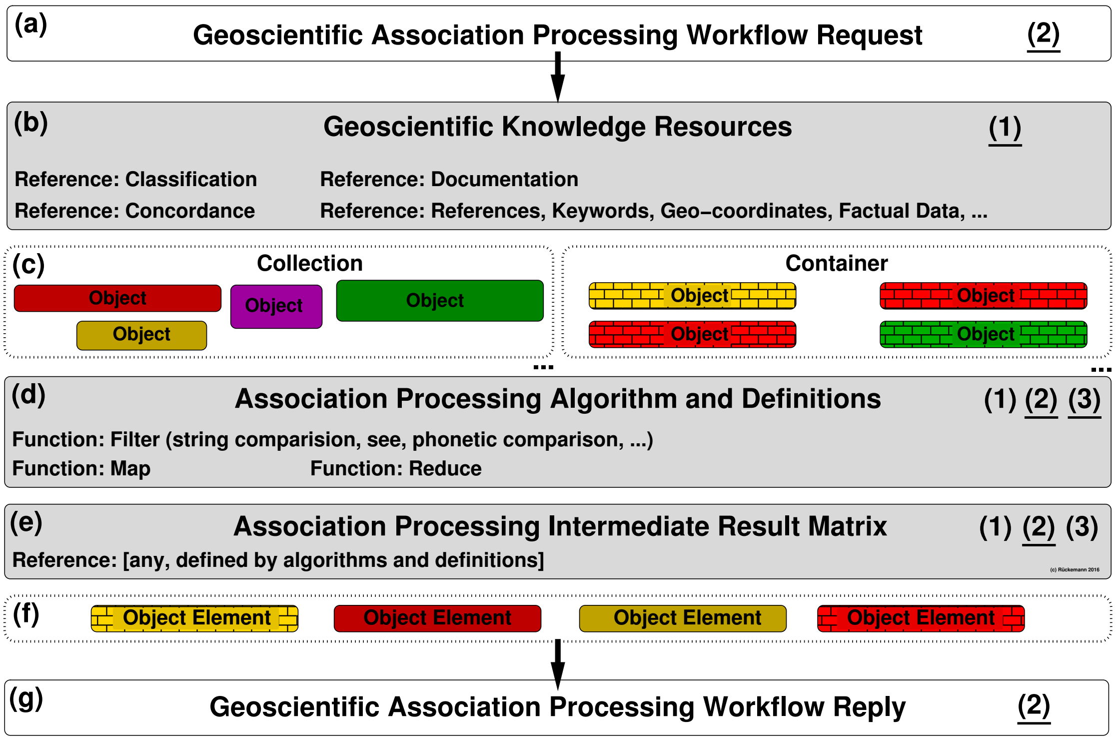

Figure 2. Geoscientific association processing workflow: Creation of intermediate result matrices from geoscientific resources and references (collections and containers) in reply to workflow requests. Workflow steps are labelled with lowercase letters. Layers are labelled with numbers, primary layers are underlined.

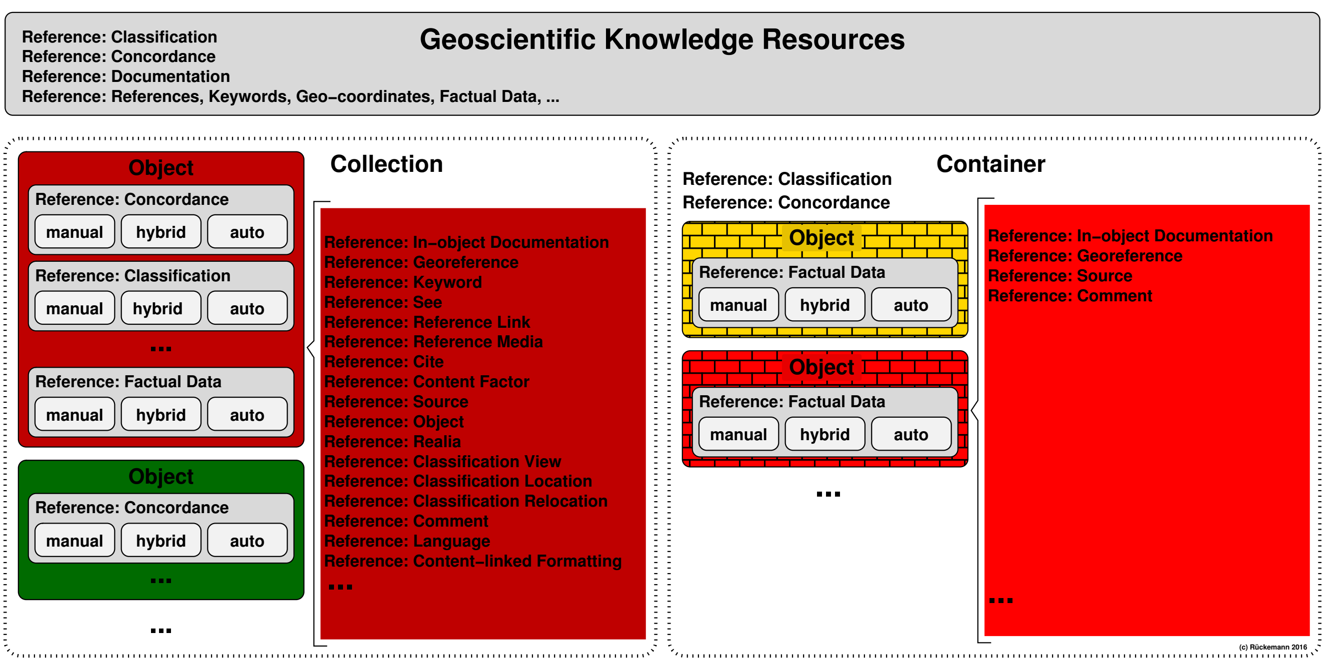

Figure 3. Geoscientific knowledge resources and objects: Selected knowledge resources' objects containing references for concordances and classifications in collections and containers. The excerpt illustrates a distinct handling of manually, hybrid, and automatically created data.

IV. Implementation of reference types

Objects can carry any type of references. Objects can be grouped, e.g., in collections or containers. When larger groups are created then also these groups can carry their references. These references may occur in any combination but in practice these references will be a subset or a complementary set to the objects' references. Objects can be created by manual, automated, and hybrid means. Therefore, any type of references of that kind may exist.

Tables I and II show excerpts of the references, which were added to be used within the knowledge resources for two types of object groups, namely collections and containers.

TABLE I. Geoscientific Knowledge Resources' Collection and Container References Types used for Processing (excerpt).

References Types	*Group and Implementation*	*Example*
Classification	O & C	UDC
Concordance	O & C	UCC
In-object documentation	O & C	Text
Factual data	O & C	Text, data
Georeference	O & C	Geocoordinates
Keyword	O & C	Text
See	O & C	Text
Reference link	O & C	URL
Reference media	O & C	Link
Citation	O & C	Cite, bib
Content Factor	O & C	CONTFACT
Source	O & C	Text
Object	O & C	Text, code
Realia	O & C	Text
View, conceptual	O & C	Classification view
Location, conceptual	O & C	Classification location
Relocation, conceptual	O & C	Classification relocation
Comment	O & C	Text
Language	O & C	EN, DE
Content-linked formatting	O & C	Markup, LaTeX

TABLE II. Geoscientific Knowledge Resources' implemented External References Types used for Processing (excerpt).

References Types	*Group and Implementation*	*Example*
Tag	E	Text, tags
Index Entry	E	Text, idx
Glossary Entry	E	Text, glo
List Entry	E	Text, lis

The reference types are organised in three major groups: Object collections (O), object containers (C), and external or externally created references (E).

This case study primarily addresses geoscientific and archaeological resources. The resources were extended for using a multitude of references types of creating associations (Tables II and I). Therefore, the resources especially contain georeferences, UDC classifications for any object, including complex conceptual knowledge, geoclassification, concordances like Universal Classified Classification (UCC) [12], and Content Factors [13] in order to describe [14] the content. Many reference types are part of the objects. Nevertheless, in practice, the organisation of references is more uniform within containers.

The reference types shown provide a lot of information regarding content and context, which could otherwise not be deducted from the object data itself. In addition, all reference types may exist in multiple views, multiple languages, and multiple context – any of which can be added in instances created by manual, automatic, and hybrid means.

V. Creation of associations and processing means

As far as the algorithms implemented in components carry essential information for processing and computation, e.g., for creating new results and output, they should be documented with the knowledge resources themselves. As associations can be created by arbitrary workflows, it is most important to know, which components can carry which facilities and how to exploit, e.g., in a multi-disciplinary context like geosciences and archaeology. Important aspects for the quality of results of the association processing are geocoordinates, associations, conceptual knowledge, and supporting methodologies.
Geocoordinates' data is for example used to spatially select and associate objects. Geocoordinates can be part of

- knowledge objects,
- containers and container objects,
- references (e.g., knowledge resources' references or Google Maps references).

Conceptual knowledge data, e.g., used for the classification of objects of any kind, can be part of

- knowledge objects,
- containers and container objects,
- unstructured data, mostly used with automated processes with lower quality results.

Associations can especially result from rich content, from constellations of content and

- context in object collections,
- context in object containers,
- in-text references (comparisons, see, ...),
- internally,
- link-references (links to external resources),
- external resources,
- between all components.

Supporting methodologies and technologies, which can be exploited for the creation and processing of associations are

- string comparisons,
- transliterations,
- phonetics,
- statistics,
- metadata,
- Content Factor,
- object elements rhythm,
- dynamical data, ...

Associations were used for developing knowledge resources, optimising result matrices, e.g., within knowledge discovery workflows, creating concordances, creating references, improving knowledge objects and resources, gaining new knowledge. The combination allows various degrees of precision and fuzzyness as required for spanning multi-disciplinary and multi-lingual data. An optimisation can improve the quality of data, especially the quality of associations introduced for automated classification of unstructured data.

VI. Exploitation of computation facilities

Within the layers, there are three kinds of facilities, which are targets to be exploited by computation.

(1) **First block:** Knowledge resources.
Storage services and resources.
- Purpose: Data.
- Task/Method: Creation, development, maintenance processes.
- Implementation: Editing components, versioning tools, and high end text editors are used together with automation tools and scripting. The knowledge resources themselves are based on fully portable structures and markup.

(2) **Second block:** Services and interfaces.
Layer: Dedicated infrastructure.
- Purpose: Workflows.
- Task/Method: Scripting, workflow description.
- Implementation: Perl, Tcl, and C are used for an implementation.

(3) **Third block:** Processing.
Layer: Compute services and resources.
- Purpose: Individual and parallelised processes and tasks as well as dynamical and interactive processes.
- Task/Method: Batch system, configuration, dedicated resources.
- Implementation: Here, Portable Batch System (PBS), Torque, and Moab are used, formerly also IBM LoadLeveler and Condor. As far as required for a certain scenario also dynamical or interactive jobs can be executed.

The exploitation of computation facilities is mostly based on these three featured component groups and the described implementation. This realises the purposes of extracting data and information, utilising workflow scripts, and submitting dynamical and batch jobs.

VII. Integration and association case study

The implementation has been done according to the described architecture and enabling the required association processing workflows based on the available components. Therefore, the major implementation tasks concentrated on the content related facilities, especially the geoscientific and archaeological knowledge resources, and the application components, which delivered an appropriate starting point for a case study. The respective features were created in the knowledge resources' objects, which were under continuous development over the last decades in the LX knowledge resources. The application components have been extended and configured to work with the required application scenario. This includes the dynamical components from the Geo Exploration and Information (GEXI) project, e.g., the actmap components, based on Perl and Tcl scripting, C and Fortran.

The implementation for the case study was integrated with the the base for this case study, the long-term knowledge resources (LX Foundation Scientific Resources), which were developed and used over several decades, including geoscientific and archaeological objects and containers. A case study example based on the created resources is presented with the following workflow.

A. Volcano and Rhodos association discovery workflow

The workflow starts with the target to find possible associations and links between "Vesuvius" and "Rhodos".

1) Entry nodes: Vesuvius – Rhodos (Rhodos/Rhodes etc.).
2) Criteria and definition set.
3) Filter association processing criteria.
4) Filter association processing.
5) Selection and generation of compute instructions.
6) Sorting.
7) Formatting.
8) Selection.
9) First level association - both nodes.
10) Second level association.
11) Common object 1 and 2 (level 1).
12) Common object 11 and 22 (level 2).

Steps 2) to 5) of the workflow analyse and implement the criteria and definitions for the request and prepare the appropriate compute instructions. Steps 6) to 8) handle the sorting, formatting, and the selection of the intermediate result matrix. Steps 9) and 10) generate a first level association and after that a second level association. The concluding steps 11) and 12) generate the common objects for levels 1 and 2.

B. Resources and content

As an example, an object excerpt for one of the entry nodes is shown in Figure 4, which shows a referenced Vesuvius collection object containing factual and conceptual knowledge.

```
Vesuvius [Volcanology, Geology, Archaeology]:
          (lat.) Mons Vesuvius.
          (ital.) Vesuvio.
          Volcano, Gulf of Naples, Italy.
          Complex volcano (compound volcano).
          Stratovolcano, large cone (Gran Cono).
          Volcano Type: Somma volcano,
          VNUM: 0101-02=,
          Summit Elevation: 1281\UD{m}. ...
          Syn.: Vesaevus, Vesevus, Vesbius, Vesvius
          s. volcano, super volcano, compound volcano
          s. also Pompeji, Herculaneum, seismology
          compare La Soufrière, Mt. Scenery, Soufriere
          %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(4+37+23+24)=12=14
          %%IML: GoogleMapsLocation: http://maps.google.de/maps?hl=de&gl=de&vpsrc
           =0&ie=UTF8&ll=40.821961,14.428868&spn=0.018804,0.028238&t=h&z=15
```

Figure 4. Workflow entry node: Knowledge resources collection object "Vesuvius" (LX resources, geoscientific collection, excerpt).

The object carries names and synonyms in different languages, dynamically usable geocoordinates, UDC classification and so on, including geoclassification (UDC:(37), Italia. Ancient Rome and Italy). The listing in Figure 5 shows an instance of a container entry excerpt from a volcanological features container.

```
CONTAINER_OBJECT_EN_ITEM: Vesuvius
CONTAINER_OBJECT_EN_PRINT: Vesuvius
CONTAINER_OBJECT_EN_COUNTRY: Italy
CONTAINER_OBJECT_EN_CONTINENT: Europe
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_XX_VNUM: 0101-02= ...
```

Figure 5. Processed instance of a simple knowledge resources container entry (LX resources, geoscientific container, excerpt).

The container component contains a large number of volcanic features and volcanoes, like Vesuvius, Thera, and Santorini. 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 common 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. The criteria and definitions are given by variables (Figure 6).

```
MATRIXLX
MATRIXRESLEV1
MATRIXRESLEV11
MATRIXRESLEV2
MATRIXRESLEV22
```

Figure 6. Criteria and definitions: Variables (LX Resources, excerpt).

The resource levels instruct the routines to execute two levels, one primary plain discovery each and a secondary in-depth discovery considering the primary results. The filter, selection, and processing instructions are handled by generators. The internal sequence is shown in Figure 7.

```
gen_matrix_pipe_level0_level1 "Vesuvius"
gen_pipe_reslevel1 | \
gen_grep_formatrix | \
gen_grep_forstrip
...
gen_matrix_pipe_level0_level1 "Rhod.s"
gen_pipe_reslevel11 | \
gen_grep_formatrix | \
gen_grep_forstrip
...
gen_matrix_pipe_level0_level2 "Kameiros"
gen_matrix_pipe_level2_level22 "Pozzolan"
gen_pipe_reslevel22 | \
gen_grep_formatrix | \
gen_grep_forstrip
```

Figure 7. Sequence of association routines for discovery workflow, dual-level (LX Resources, excerpt).

The sort, formatting, and selection are done with the function calls (`forstrip`). The "Vesuvius – Rhodos" association delivers "Kameiros", "Thera", "Santorini", and further intermediate result matrix elements from the secondary in-depth discovery.

C. The Kameiros' material results

The case study integrates the geoscientific and archaeological collection and container context and English entries. Figure 8 shows an excerpt of a referenced Kameiros object entry with UDC classification, media, and citation references, including geoclassification (UDC:(38), Ancient Greece).

```
Kameiros [Archaeology, Geophysics, Remote Sensing, Seafaring]:
         Greek city, Rhodos Island, Dodekanese, Greece.
         Modern location name Kámiros, Greece.
         ...
         Object:             Ancient architecture, stone, cement.
         Object-Keywords:    water tank, cement, lower area
         Object-Type:        Realia object.
         Object-Location:    Kameiros, Rhodos, Greece.
         Object-FindDate:    2011-10-27
         Object-Photo:       Claus-Peter Rückemann, ...
         %%SRC: 2013 CPR
         %%IML: media: YES 20130922 {LXC:DETAIL----} {UDC:(0.034)(38)770}
          LXDATASTORAGE://.../img_1342.jpg
         %%IML: UDC-Object:[902+903.2]+691.54+720.32+(38)+(4)
         ...
         %%IML: cite: YES 19980000 {LXK: concrete; pozzolan; Kameiros; Rhodos;
          Rhodes; Greece; Archaeology; Geosciences} {UDC:...} LXCITE://
          Koui:1998:Kamirian
         %%IML: cite: keyword: object: water storage tank
         %%IML: cite: keyword: material: concrete; Santorine earth mixed;
          natural cement; volcanic earth; lime
         %%IML: cite: keyword: location: Kameiros; Kamiros; Rhodos; Rhodes;
          Thera; Santorine; island of Yali; island of Nisyros
         ...
         %%IML: cite: YES 20120000 {LXK: cement; pozzolan; Kameiros; Rhodos;
          Rhodes; Greece; Archaeology; Geosciences} {UDC:...} LXCITE://
          Snellings:2012:Cementitious
         ...
         %%IML: cite: YES 20110000 {LXK: concrete; pozzolan; Kameiros; Rhodos;
          Rhodes; Greece; Archaeology; Geosciences} {UDC:...} LXCITE://
          Courland:2011:Concrete
         ...
         %%IML: cite: YES 20110000 {LXK: Archaeology; Geosciences; Vesuvius;
          Pompeji} {UDC:...} LXCITE://Hartge:2009:Vesuvius
         vgl. Rhodos, Iálissos, Lindos, Akandia
```

Figure 8. Association result matrix element, object "Kameiros" (LX resources, archaeological collection, excerpt).

The association processing "Vesuvius – Rhodos" revealed the reference to Vesuvius / (via Kameiros-associated citations) Pozzuoli / pozzolan. The excerpt also delivers a number of associated references on ancient concrete technology [15], cementitious materials [16], history of concrete [17], and evolution of concrete [18].

Looking for secondary documentation on eruptions being associated with Pozzuoli, e.g., the 1631 eruption of Vesuvius, delivers bibliographic sources like [19], which provides a lot of unique context information from an original source. This means there are several associations linking Vesuvius with Rhodos and one link is a technology, based on material from geoscientific context, documented in an archaeological site.

Figure 9 shows an excerpt of a referenced Thera object entry with UDC classified knowledge objects delivered from the association processing.

```
Thera    [Volcanology, Geology, Archaeology]:
         Volcano, \lxidx{Santorini}, Greece.
         Volcano Type: Somma volcano,
         Country: Italy,
         Subregion Name: Italy,
         VNUM: 0101-02=,
         VEI: VEI7,
         Volcano Status: Historical,
         Summit Elevation: 564\UD{m}.
         s. VEI
         %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(4+38+23+24)=14
         %%IML: keyword: volcano, eruption, Santorini, Thera, crete, Minoan
          civilisation, culture, Mycenae, culture, fleet, volcanic ash, vanish,
          rise, historical city, catastrophe
```

Figure 9. Association result matrix element, object "Thera" (LX resources, excerpt).

The matrix element delivers additional information and references, e.g., to volcanic features containers via the volcano number (VNUM) and the Volcanic Explosivity Index (VEI).

In this means, the knowledge resources provide data and references for the association routines. Result matrices are computed from the association result matrix elements (example excerpts Figures 8 and 9), which have been deducted from the knowledge resources using the association routines.

The above sequence of association routines (Figure 7) was used for the creation of a result matrix (routines implemented in `lxgrep_in_depth`). The listing in Figure 10 shows an extended excerpt of the intermediate result matrix output for this case example.

```
MATRIXentry{Vesuvius}
MATRIXcitekeywords{location: Vesuvius, Italy}
MATRIXindex{pozzolan}
MATRIXindex{Campi Flegrei}
MATRIXindex{Pozzolana}
MATRIXindex{Pozzuoli}
MATRIXindex{Puteoli}
MATRIXkeywordcontext{keyword-Context: KYW :: 1634-1676 Polyhistor ...}
MATRIXkeywordcontext{keyword-Context: KYW EN :: earthquake, seismology, seismics
 , Fogel, Fogelius, volcano, Vesuvius}
MATRIXkeywordcontext{keyword-Context: KYW REF S. 62 :: Vesuvius; pyroklastischer
  Strom; Aschewolke; Pozzuolo; Dreißigjähriger Krieg}
MATRIXkeywordcontext{keyword-Context: KYW REF S. 63 :: Schwefel, Solfatara,
 Fumarole}
MATRIXkeywordcontext{keyword-Context: KYW :: terrae motus, terra motus, motus
 terrae, Erdbeben, Seismologie, Seismik, Erdbewegungen}
MATRIXkeywordcontext{keyword-Context: KYW :: Vulkanismus, Vulkanologie, Vesuv,
 Vesuvius, Vesuvium, Erdbeben, Beben}
MATRIXkeywordcontext{keyword-Context: TXT :: 1631/1632 16xx, terra motus,
 fogelius}
MATRIXkeywordcontext{keyword-Context: TXT :: Fogelius, Historici Pragmatici
 universal, Terrae motus, Physica}
MATRIXkeywordcontext{keyword-Context: TXT REF S. 175 :: Pozzuolo ...}
MATRIXkeywordcontext{keyword-Context: TXT REF S. 183 :: Pozzo di Somma}
MATRIXkeywordcontext{keyword-Context: TXT REF S. 73 :: Grotta del Cane}
MATRIXseealso{phlegra, Solfatara}
MATRIXseealso{Traß}
MATRIXsynonym{Puzzuolana}
MATRIXsynonym{Vesaevus, Vesevus, Vesbius, Vesvius}
...
MATRIXentry{pozzolan}
MATRIXindex{diatomaceous earth}
MATRIXindex{iron oxide}
MATRIXindex{Kameiros}
MATRIXindex{Kamiros}
MATRIXindex{Phlegraean Fields}
MATRIXindex{Pozzolana}
MATRIXindex{pozzolanic activity}
MATRIXindex{pozzolanic earths}
MATRIXindex{pozzolanic material}
MATRIXindex{Pozzuoli}
MATRIXindex{Puteoli}
MATRIXindex{Rhodes}
MATRIXindex{tuffs}
MATRIXindex{Vesuvius}
MATRIXseealso{Traß}
MATRIXsee{Puzzolane}
MATRIXsynonym{Puzzuolana}
...
MATRIXentry{Kameiros}
MATRIXcitekeywords{material: concrete; Santorine earth mixed; natural cement;
 volcanic earth; lime}
MATRIXcitekeywords{material: pozzolan}
MATRIXcitekeywords{material: Pozzuoli volcanic pumice}
MATRIXcitekeywords{material: stone called Santorini}
MATRIXcitekeywords{method: Ternary CaO-SiO2-Al2O3 diagram, major SCM groups}
MATRIXcitekeywords{object: water storage tank}
MATRIXcompare{Rhodos, Iálissos, Lindos, Akandia}
MATRIXindex{Pozzolana}
MATRIXindex{Pozzuoli}
MATRIXindex{Puteoli}
MATRIXindex{Puzzolana}
MATRIXindex{Vesuvius}
MATRIXobjectkeywords{Object-Keywords: water cistern, top area}
MATRIXobjectkeywords{Object-Keywords: water pipeline, clay, upper area}
MATRIXobjectkeywords{Object-Keywords: water tank, cement, lower area}
MATRIXseealso{Traß}
MATRIXsee{pozzolan}
MATRIXsee{Puzzolane}
MATRIXsee{Rhodos}
MATRIXsynonym{Puzzuolana}
MATRIXsynonym{Puzzuolane}
MATRIXtextintext{Kámiros, Greece}
MATRIXtextintext{Object: Ancient architecture, stone, cement}
MATRIXtextintext{Object-Location: Kameiros, Rhodos, Greece}
MATRIXtextintext{Pozzolan}
MATRIXtextintext{Puzzolane}
```

Figure 10. Intermediate result matrix output, groups (excerpt).

The example contains excerpts of three different element groups. The according object entries are “Vesuvius”, “pozzolan”, and “Kameiros”.

The potential number of referenced objects in an intermediate result matrix depends on the references sources and the discovery algorithms. The number used for further steps depends on selection and filter criteria applied for the respective context. The potential elements and the number of elements in a group can vary depending on the resources and objects referred.

The groups result from association processing of knowledge resources’ objects. The groups contain intermediate matrix element references referring to the objects integrated into the knowledge resources. Examples for this case are shown in Table III.

TABLE III. INTERMEDIATE MATRIX ELEMENT REFERENCES FROM ASSOCIATION PROCESSING OF KNOWLEDGE OBJECTS (EXCERPT).

Matrix Element	*Meaning*
`MATRIXentry`	Object entries in knowledge res.
`MATRIXcitekeywords`	Citation keywords
`MATRIXindex`	Index entries
`MATRIXkeywordcontext`	Context keywords
`MATRIXsee`	‘see’ references
`MATRIXseealso`	‘see also’ references
`MATRIXsynonym`	Synonym references
`MATRIXcompare`	‘compare’ references
`MATRIXobjectkeywords`	Object keywords
`MATRIXtextintext`	References of text in object text

The matrix elements allow various precise (e.g., see, see also) and fuzzy (e.g., keywords) references, which are used for association processing algorithms.

When we extend the discovery and integrate chronological and associated objects and locations from the resources then the result matrix also includes years with volcanic, geological, geophysical, and technological context. Possible sources are collections and containers.

The listing in Figure 11 shows a representation of additional result matrix elements associated for this case after these attributes were considered.

```
MATRIXtextintext{date: -300000 Vesuvius, volcanic activity, oldest deposits}
MATRIXtextintext{date: -001800 Vesuvius, volcanic activity, Avellino eruption}
MATRIXtextintext{date: -001680 Santorin, Aegean, volcanic eruption, Thera}
MATRIXtextintext{date: -000700 Vesuvius, volcanic activity}
MATRIXtextintext{date: -000227 Rhodos, seismic activity}
MATRIXtextintext{date:  000062 Vesuvius, seismic activity, earthquake, Pompeji
 destruction}
MATRIXtextintext{date:  000079 Vesuvius, volcanic activity, explosive eruption,
 ash cloud, tuff, Pompeji destruction, Herculaneum, Stabiae}
MATRIXtextintext{date:  000142 Rhodos, seismic activity}
MATRIXtextintext{date:  000202 Vesuvius, volcanic activity}
MATRIXtextintext{date:  000345 Rhodos, seismic activity}
MATRIXtextintext{date:  000472 Vesuvius, volcanic activity}
MATRIXtextintext{date:  000512 Vesuvius, volcanic activity}
MATRIXtextintext{date:  000515 Rhodos, seismic activity}
...
MATRIXtextintext{location: Kameiros, island Rhodes, Greece; Kamiros, Greece;
 Rhodos; Rhodes}
MATRIXtextintext{location: Thera; Santorine; island Yali; island Nisyros}
MATRIXtextintext{location: Vesuvius}
MATRIXtextintext{location: Solfatara, Vesuvius}
MATRIXtextintext{location: Pantheon, Rome}
MATRIXtextintext{location: Caesarea Maritima}
MATRIXtextintext{location: Hagia Sophia, Konstantinopel}
...
MATRIXtextintext{material: pozzolan}
MATRIXtextintext{material: volcanic tuff}
MATRIXtextintext{material: Opus caementitium}
MATRIXtextintext{material: pozzolanic activity}
MATRIXtextintext{material: pozzolanic earths}
MATRIXtextintext{material: pozzolanic material}
MATRIXtextintext{material: volcanic ashes}
MATRIXtextintext{material: diatomaceous earth}
```

Figure 11. Additional result matrix entries for intermediate result matrix associated with integrated resources (excerpt).

The example contains excerpts of additional entries resulting for the three according object entries "Vesuvius", "pozzolan", and "Kameiros" from the above example case. Two aspects are most interesting for the discovery:

1) The new content and context, e.g., the context of Pantheon, Caesarea Maritima, and the Hagia Sophia in the context of Kameiros, Rhodos.
2) The possibility of referencing classifications to groups of content and context, e.g., date, location, and material.

These groups contain additional matrix elements referring to the objects integrated into the knowledge resources. An excerpt of additional matrix entry references from objects is shown in Table IV.

TABLE IV. ADDITIONAL INTERMEDIATE MATRIX ELEMENT REFERENCES FROM ASSOCIATION PROCESSING (EXCERPT).

Matrix Element	*Meaning*
`MATRIXtextintext...date: ...`	Date reference
`MATRIXtextintext...location: ...`	Location reference
`MATRIXtextintext...material: ...`	Material reference

In this case, all the additional references result from texts integrated in object texts (`MATRIXtextintext`). The associated additional information from the object references contains chronological information, location information, and material information.

The overall result is a very rich matrix. With its elements, an advanced matrix links different content and context from many hundreds to many thousands of objects and sources. The listing depicts the content of the result matrix in a readable formatting and excerpts some elements.

The matrix also contains references to the source data within the knowledge resources (for examples on media and citation references see Figure 8) and also refers to many other structured and unstructured data, e.g.,

- terms,
- names,
- locations,
- georeferences,
- bibliographic data,
- citations,
- classification, and
- media data.

As has been shown, media data can be integrated and documented with the knowledge resources, e.g., either as objects or sub-objects. This documentation also allows dependencies between object.

D. The integrated Kameiros' media references results

Media objects contain own references, e.g., classification, citations, documentation, and keywords and can therefore contribute in many ways to new insight – besides their intrinsic media content. The following photo data (Figure 12) from the media references for "Kameiros" was delivered by the result matrix.

The photos have been taken in 2013 by the Knowledge in Motion (KiM) natural sciences and archaeology sections in the ancient city of Kameiros on Rhodos, Greece, during the Geo Exploration and Information (GEXI) Eastern Mediterranean research and studies campaign. Today, the location on the western coast of the island of Rhodes is named Kamíros.

Figure 12. Integrated media photo objects associated with the knowledge object "Kameiros", referring to pozzolane and Vesuvius (LX resources).

The data shows the respective photo objects integrated with the knowledge resources and usable with the association processing, which were resulting for the presented case study. The sort order of the photo object chain reflects the arrangements of the locations: On top level to bottom level of the excavated site (photo objects from left to right, from top to bottom). In detail, the results show in this sort order

1) the top level water tank,
2) the top level water pipeline following downhill,
3) an excavated part of the water pipeline,
4) a lower level water tank,
5) a single element of the water pipeline system,
6) and a bottom level water tank.

Object chains can be created by their objects' attributes, e.g., in this case by classification, spatial position, and altitude.

The water tanks are coated with the 'pozzolan' cement material, part of which are still in place, e.g., in the bottom level water tank.

Each of the objects in the object chain contains classification references to the pozzolan material, therefore the objects provide the missing link: These references from ancient Kameiros are also associated with Vesuvius volcano and refer to the later Roman adoption of comparable cement 'pozzolan' technology. These Rhodian realia are widely associated with natural phenomena, therefore this is reflected for the objects by many references [20] to respective phenomena and other places, e.g., on Rhodes, like Epta Piges, Rodini Park, the Nymphaion of Rhodes, cisterns, as well as geological water level marks.

Continuation of the case study [21] has conceived the documentation available and planning the additional research and development and the data to be collected and added to the knowledge resources.

VIII. Evaluation

The integration of multi-disciplinary knowledge resources and association processing can provide large benefits for creating result matrices and content. The integration and implementation is consequently data-centric. The implementation is as well as far data-centric as possible: "The term data-centric refers to a focus, in which data is most relevant in context with a purpose. Data structuring, data shaping, and long-term aspects are important concerns. Data-centricity concentrates on data-based content and is beneficial for information and knowledge and for emphasizing their value. The value of knowledge depends on the 'currency' used for individual scenarios.

The value of an integration depends on the scenario and priorities. There is no general knowledge-based means available for defining the value of integration and implementation other than by the value of knowledge gained. Technical implementations need to consider distributed data, non-distributed data, and data locality and enable advanced data handling and analysis. Implementations should support separating data from technical implementations as far as possible." [22].

According to the definition this means, that data and technical implementations can be separated and the created knowledge resources and technical components comply to the above criteria.

The structure and the aggregation of references increase the flexibility of possible workflows. Increasing the quality of data in the described type of long-term knowledge resources – by including references – can increase the quality of result matrices from discovery processes.

The examined case showed that a technology and material, which have not been explicitely documented in context of a knowledge resources object, can be associated with the context of different objects. Here, the Greek origin of the "pozzolan" technology was associated, which was named after the later use in Roman times.

Association processing can support discovery processes even when references are not explicitely available in text and documentation, and would therefore be unexpected or unknown. Association processing can use multi-level discovery in order to gain additional information, which is not visible from an otherwise isolated documentation.

The developed structures and methods can be widely beneficial for knowledge development and discovery as well as for creating implementations for advanced discovery components. The methodology allows to extend and exploit long-term multi-disciplinary content documentation and discovery and gain new insights from otherwise not associated data.

IX. Conclusion

This paper presented the research on the integration of knowledge resources and advanced association processing and presented features and benefits for the processing of associations. Major results from the resources side as well as from the research on geosciences and archaeology were shown.

First, the research shows that knowledge resources and association processing can be integrated most efficiently and flexibly. The major reason is, that structured knowledge resources can be successfully extended for allowing a multitude of integrated references types, e.g., geoclassification and media. The structured resources can also carry references and associations to differently structured or unstructured resources. Besides structured resources, also scenarios based on unstructured resources can benefit from an integration.

Second, the implementation shows that integration is least invasive to the knowledge resources and to the workflows. For both, knowledge resources and implementations, an arbitrary individual development can be preserved. The required parts of the workflows can stay modular.

The elements from associations contained in the result matrix are not procurable when using only plain methods like simple string search or plain discovery. Furthermore, the integration of methods, e.g., association, classification, and phonetic algorithms, allows any degree of precision and fuzziness.

A layered concept can be deployed with the implementation of computation and storage facilities. From the structural and knowledge point of view, the extended features are least invasive to the described type of knowledge resources and procedures.

From the geoscientific and archaeological side the factual results are most notable because the methodology integrates multi-disciplinary and multi-lingual knowledge beyond conventional means and shows a large number of associations, which cross multiple disciplines and languages.

The association processing has shown to benefit from additional elements referring to objects added to intermediate result matrices. Different data and media can be included, e.g., by documentation and classification, which vastly extends the spectrum of possible applications and can foster new insight.

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 further developing the multi-disciplinary knowledge resources and creating advanced methods for describing the content and context of objects.

It will be interesting to see applications of an integration of knowledge resources and association processing as resulting from this research. Feedback on the used 'currency' and the value of knowledge and data from different scenarios will be welcome. The new method should carry facilities for supporting long-term knowledge development and analysis as well as for enabling automation and High End Computing.

Acknowledgements

We are grateful to the "Knowledge in Motion" (KiM) long-term project, Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), for partially funding this implementation, case study, research and studies campaign, and publication under grant D2012F1P04497 and to its senior scientific members, especially to Dr. Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek (GWLB) Hannover, to Dipl.-Biol. Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, and to Dipl.-Ing. Martin Hofmeister, Hannover, for the contributions to the research on the development and application of classifications. and for fruitful discussion, inspiration, practical multi-disciplinary case studies, and the analysis of advanced concepts. We are grateful to Dipl.-Ing. Hans-Günther Müller, Cray, for his work on flexible practical solutions to architectural challenges and excellent technical support. We are grateful to all national and international partners in the Geo Exploration and Information cooperations for their constructive and trans-disciplinary support. We thank the Science and High Performance Supercomputing Centre (SHPSC) for long-term support of collaborative research since 1997, including the GEXI developments and case studies.

References

[1] C.-P. Rückemann, "Advanced Association Processing and Computation Facilities for Geoscientific and Archaeological Knowledge Resources Components," in Proceedings of The Eighth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2016), April 24–28, 2016, Venice, Italy. XPS Press, 2016, Rückemann, C.-P. and Doytsher, Y. (eds.), pages 69–75, ISSN: 2308-393X, ISBN-13: 978-1-61208-469-5, ISBN-13: 978-1-61208-060-4 (CDROM), TMDL: geoprocessing_2016_4_20_30144, URL: http://www.thinkmind.org/download.php?articleid=geoprocessing_2016_4_20_30144 [accessed: 2016-06-05], URL: http://www.thinkmind.org/index.php?view=article&articleid=geoprocessing_2016_4_10_30144 [accessed: 2016-06-05].

[2] C.-P. Rückemann, "Integrated Computational and Conceptual Solutions for Complex Environmental Information Management," in The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences, Proceedings of The 13th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 23–29, 2015, Rhodes, Greece, Proceedings of the American Institute of Physics (AIP), vol. 1738. AIP Press, 2016, Simos, T. E., Tsitouras, C. (eds.), ISBN-13: 978-0-7354-1392-4, ISSN: 0094-243X (American Institute of Physics Conference Proceedings, print), ISSN: 1551-7616 (online) (eISSN), DOI: 10.1063/1.4951833.

[3] "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: 2016-01-10].

[4] "Tcl Developer Site," 2016, URL: http://dev.scriptics.com/ [accessed: 2016-01-10].

[5] "ISO 14000 - Environmental management," 2016, URL: http://www.iso.org/iso/iso14000 [accessed: 2016-01-10].

[6] C.-P. Rückemann, "Enabling Dynamical Use of Integrated Systems and Scientific Supercomputing Resources for Archaeological Information Systems," in Proceedings INFOCOMP 2012, Oct. 21–26, 2012, Venice, Italy, 2012, pp. 36–41, ISBN: 978-1-61208-226-4, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2012_3_10_10012 [accessed: 2016-08-28].

[7] "LX-Project," 2016, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX [accessed: 2016-06-05].

[8] D. T. Meridou, U. Inden, C.-P. Rückemann, C. Z. Patrikakis, D.-T. I. Kaklamani, and I. S. Venieris, "Ontology-based, Multi-agent Support of Production Management," in The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences, Proceedings of The 13th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 23–29, 2015, Rhodes, Greece, Proceedings of the American Institute of Physics (AIP), vol. 1738. AIP Press, 2016, Simos, T. E., Tsitouras, C. (eds.), ISBN-13: 978-0-7354-1392-4, ISSN: 0094-243X (American Institute of Physics Conference Proceedings, print), ISSN: 1551-7616 (online) (eISSN), DOI: 10.1063/1.4951834.

[9] UDC, Universal Decimal Classification. British Standards Institute (BSI), 2005, Complete Edition, ISBN: 0-580-45482-7, Vol. 1 and 2.

[10] "UDC Online," 2016, URL: http://www.udc-hub.com/ [accessed: 2016-01-10].

[11] C.-P. Rückemann, "From Multi-disciplinary Knowledge Objects to Universal Knowledge Dimensions: Creating Computational Views," International Journal On Advances in Intelligent Systems, vol. 7, no. 3&4, 2014, pp. 385–401, ISSN: 1942-2679, URL: http://www.iariajournals.org/intelligent_systems/intsys_v7_n34_2014_paged.pdf [accessed: 2016-08-28].

[12] C.-P. Rückemann, "Creation of Objects and Concordances for Knowledge Processing and Advanced Computing," in Proceedings of The Fifth International Conference on Advanced Communications and Computation (INFOCOMP 2015), June 21–26, 2015, Brussels, Belgium. XPS Press, 2015, pp. 91–98, ISSN: 2308-3484, ISBN-13: 978-1-61208-416-9, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2015_4_30_60038 [accessed: 2016-08-28].

[13] C.-P. Rückemann, "Enhancement of Knowledge Resources and Discovery by Computation of Content Factors," in Proceedings of The Sixth International Conference on Advanced Communications and Computation (INFOCOMP 2016), May 22–26, 2016, Valencia, Spain. XPS Press, 2016, Rückemann, C.-P., Pankowska, M. (eds.), pages 24–31, ISSN: 2308-3484, ISBN-13: 978-1-61208-478-7, ISBN-13: 978-1-61208-061-1 (CDROM), TMDL: infocomp_2016_2_30_60047, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2016_2_30_60047 [accessed: 2016-06-18], URL: http://www.thinkmind.org/index.php?view=article&articleid=infocomp_2016_2_30_60047 [accessed: 2016-06-18].

[14] F. Hülsmann, C.-P. Rückemann, M. Hofmeister, M. Lorenzen, O. Lau, and M. Tasche, "Application Scenarios for the Content Factor Method in Libraries, Natural Sciences and Archaeology, Statics, Architecture, Risk Coverage, Technology, and Material Sciences," KiM Strategy Summit, March 17, 2016, Knowledge in Motion, "Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF)", Hannover, Germany, 2016.

[15] M. Koui and C. Ftikos, "The ancient Kamirian water storage tank: A proof of concrete technology and durability for three millenniums," Materials and Structures, Nov. 1998, Vol. 31, Issue 9, 623–627, DOI: 10.1007/BF02480613, ISSN: 1359-5997 (print), ISSN: 1871-6873 (online), Kluwer Academic Publishers.

[16] R. Snellings, G. Mertens, and J. Elsen, "Supplementary Cementitious Materials," Reviews in Mineralogy & Geochemistry, vol. 74, 2012, pp. 211–278, ISSN: 1529-6466, DOI: 10.2138/rmg.2012.74.6.

[17] R. Courland, Concrete Planet: The Strange and Fascinating Story of the World's Most Common Man-Made Material. Prometheus Books, 2011, ISBN: 978-1-61614-482-1, 396 p.

[18] A. A. Camões and R. M. Ferreira, "Technological evolution of concrete: from ancient times to ultra high-performance concrete," Structures and Architecture, 2010, pp. 1571–1578, ISBN: 978-0-415-49249-2.

[19] R. Hartge, Vesuvius, . . ., An Eye Witness Report of the Vesuvius Eruption 1631/1632 (in German: Vesuvius, Kulturgeschichtliche Betrachtungen zu den vulkanischen Kräften und Der aufgeschobene Weltuntergang? Ein Augenzeugenbericht vom Vesuv-Ausbruch 1631/1632). Verlag Die Blaue Eule, Essen, 2009, ISBN: 978-3-89924-233-1, 292 p.

[20] B. Gersbeck-Schierholz and C.-P. Rückemann, "From Past to Present: Information on Water, Wind, and Sun and Their Significance," KiM Sky Summit, Knowledge in Motion, September 25, 2016, Sky Summit Meeting, "Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF)", Hungaria, 2016.

[21] C.-P. Rückemann and B. Gersbeck-Schierholz, "Ancient Water Systems and Ageless Knowledge and Technologies," KiM Sky Summit, Knowledge in Motion, September 19, 2013, Sky Summit Meeting, "Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF)", Rhodos, Greece, 2013.

[22] C.-P. Rückemann, Z. Kovacheva, L. Schubert, I. Lishchuk, B. Gersbeck-Schierholz, and F. Hülsmann, Best Practice and Definitions of Data-centric and Big Data – Science, Society, Law, Industry, and Engineering. Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Data-centric and Big Data – Science, Society, Law, Industry, and Engineering, September 19, 2016, The Sixth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 14th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 19–25, 2016, Rhodes, Greece, 2016, URL: http://www.user.uni-hannover.de/cpr/x/publ/2016/delegatessummit2016/rueckemann_icnaam2016_summit_summary.pdf [accessed: 2016-11-06].