Advanced Scientific Computing and Multi-Disciplinary Documentation for Geosciences and Archaeology Information

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 results from the implementation of Integrated Information and Computing System components based on a new universal framework using multi-disciplinary documentation and advanced scientific computing for enabling long-term use of information in geosciences and archaeology. The solution enables the creation of general long-term knowledge resources, which will stay highly efficient, most flexible, extensible, and economic for any kind of use. The core is an architecture consisting of a knowledge resources base, application resources, and sources. The knowledge resources can contain and refer to any object. For providing a universal integration of objects and in order to improve the quality of the data, a long-term integration of structured objects and universal classification has been developed. This way, any content and context can be described and used for advanced application scenarios, deploying high end compute and storage resources for any processing and computing tasks. Due to a flexible collaboration framework the implementation allows dynamical use with information systems and supercomputing resources for any kind of workflow and application scenario, integrating knowledge from natural sciences and humanities.**

Keywords–Integrated Systems; Information Systems; Documentation; Advanced Scientific Computing; Classification; Archaeology; Geosciences; High Performance Computing.

I. Introduction

With common collections of information and data today, even if widely accessible, we are missing reliability, validation, and long-term sustainability. This is resulting from principle problems of implementation: There are no foundations of a suitable long-term strategy, documentation, tools, and resources. This demands static and dynamical components in all parts of an implementation. Content has to be developed for long periods of time. This includes new research, including historical information, and extending multi-disciplinary references. The focus question for documentation, operating on information and computing is: How can complex systems be built, developed, and extended over the necessarily long periods of time? With the application components, e.g., information system components like databases, mostly form monolithic and even proprietary blocks. Their life cycle is mostly much shorter than long-term content development. On the side of computing challenges it is possible to create solutions for very perishable present resources. Any more complex problem cannot be considered "solved" for future architectures and applications. Many information and knowledge resources cannot be used without the original context, e.g., computing resources any more. Up to now context of information science cannot be described by common means to a reasonable extend. This leads to another essential question: What information and knowledge on content and context can be preserved for medium- and long-term usage when the complexity of an overall system will be unpredictably high? The long-term strategy created here is based on an implementation architecture, which includes long-term knowledge resources with the resources and development. In this paper we concentrate on the archaeological and geosciences topics being part of the knowledge resources. The foundations should enable the essential processing of archaeological, geoscientific, geophysical, geological, spatial and other data as well as a thorough documentation of all aspects of content and context and the exploitation of advanced scientific computing methods and resources for maximum flexibility. This paper will describe the potential of Integrated Information and Computing Systems (IICS) being based on multi-disciplinary documentation for this purpose. Anyhow, it can only describe a tiny fraction of the multitude of possible features.

This paper is organised as follows. Section two describes the motivation and Section three lists aspects of related work. Section four introduces the architecture and implementation for the IICS created. Section five describes the aspects of the long-term strategy. Section six discusses an implementation case study and basic mechanisms for a geoscience-archaeology IICS. Section seven presents an evaluation including processing, computing, and classification aspects. Section eight summarises the conclusions and future work.

II. Motivation

There are no frameworks providing the necessary concepts and features for integration of data, workflows, knowledge and computing resources, and operation. For example, a lot of advanced geoscientific processing cannot be reproduced after a few years, even if the data and results are still

Copyright (c) IARIA, 2013.

existing. Therefore means for extended long-term interpretation and analysis are missing. It is a huge challenge that, besides data creation not being able to support sufficiently comprehensive documentation, the widely used technology, e.g., document formats, Uniform Resource Locators (URL), and Web Services are not persistent over longer periods of time, e.g., for static objects file formats do change, for applications the implementations will change, and for services the features will be modified. Therefore information structures built from such technologies will become inaccessible. Long-term knowledge creation cannot rely on this. From the complex systems' point of view any of those building elements are not suitable for describing objects and creating long-term knowledge resources. Anyhow the original sources and building elements are needed for documentation of the original content and context. Therefore knowledge creation has to separate the essentials of knowledge from technology, resources, and other tools while at the same time respecting their importance. Even worse , that workflows, algorithms, resources and their management cannot be guaranteed for long-term availability. The topic is very complex and experiences with long-term knowledge creation are out of scale of the time interval of most researchers. Especially, it has been found to be less difficult for groups with a strong background of classical academic education to understand the problem itself, than it is for groups with a "technical-only" background to realise the multiple benefits of classification. A basic example of long-term creation of knowledge and the implementation of applications building on these resources are presented in the following sections. It shows a scenario that can hardly be managed with other available methods and concepts in a comparable way.

III. Related Work

There is no wider concept and implementation known comparable to the solution presented, described and implemented here. Nevertheless, there are concepts for components, implementations, and terminology. Previous work [1], [2], [3], [4], [5] has delivered important basic concepts and components, e.g., Integrated Information and Computing Systems, dynamical components, and taxonomy. Taxonomy is the science and practice of classification. An important facetted classification is the Universal Decimal Classification (UDC) [6]. According to Wikipedia currently about 150000 institutions, mostly libraries, are using basic UDC classification worldwide [7], e.g., with documentation of their resources, library content, bibliographic purposes, for digital and realia objects. This is mostly restricted to publications and references but not of general knowledge and applications. Some aspects can be studied from the goals of knowledge discovery [8], which is becoming increasingly important. Other aspects are handled with search algorithms, which currently are still very primitive regarding knowledge creation and usage.

IV. Architecture and Implementation

As far, it is not commonly possible to treasure content currently used for being preserved in order to create really long-term usable content. Even much more difficult that an implementable solution for any form of long-term context is even in wide distance. In general, only a very small percentage of disciplines and researchers are familiar with knowledge classification and applications. The more, multidisciplinary classification is currently only in the focus of third parties. Operational resources and features are considered to be short-term issues whereas information, knowledge, and respective resources and features must be considered of long-term significance. Case studies showed that long-term development requires a strong sustainability of content, context, and computation. This means, the most important part of these systems is the knowledge resources containing the content and context documentation to any extent necessary for describing the activities and isolating the perishable components for context documentation. A solid classification cannot be done automatically. The more, it cannot be done automatically for use with IICS. Anyhow, in fact that different views are possible, it is reasonable to have classification view from the origin, from main disciplines or from the developers in order to increase the quality of references. The architecture respects these conditions. The following sections explain how a successful implementation of an integrated system can be created and operated using knowledge resources and classification for information system usage.

A. Architecture for documentation and development

The architecture implemented for an economical long-term strategy is based on different development blocks. Figure 1 shows the three main columns: Applications resources, knowledge resources, and originary resources.

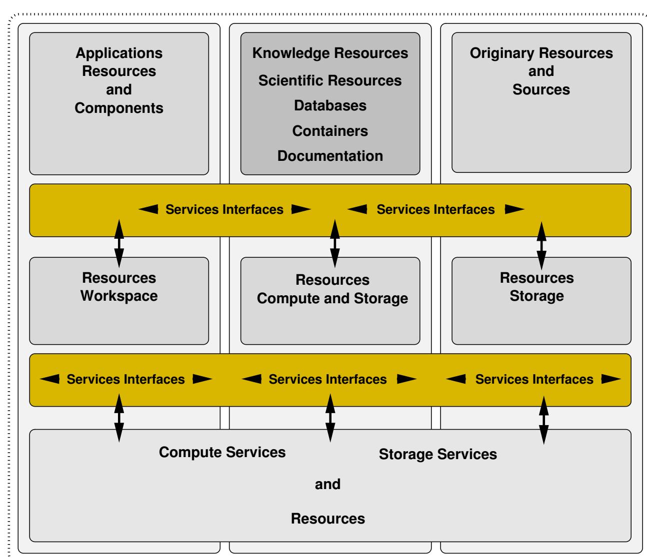


Figure 1. Architecture: Columns of practical dimensions.

The central block in the "Collaboration house" framework architecture [5], are the knowledge resources, scientific resources, databases, containers, and documentation (e.g., LX [1], databases, containers, list resources). These can be based on and refer to the originary resources and sources (photos, scientific data, literature). 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. These 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 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.

B. Components: Applications, knowledge, and sources

The main information, data, geo-referencing, and algorithms for all presented components and examples are provided by the LX Foundation Scientific Resources [1]. This deploys the structure and classification of objects necessary for a reasonable implementation. Besides the LX structure the already established Universal Decimal Classification (UDC) [6] has been integrated for objects [2] as it provides a hierarchical and multi-lingual, faceted classification for any topic and allows implementing a faceted analysis with enumerative scheme features, as well as to create new classes by using relations and grouping. In multi-disciplinary object context, this empowers to use workflows combining keywords, enumerative concepts and full-text analysis with a faceted analysis. Besides the academic, industrial, and business application scenarios in focus of the GEXI collaborations' case studies [3] it is an important factor to integrate the necessary documentation and computing facilities with systems like an Universal IICS (UIICS). An implementation of interfaces for using structure and classification with appropriate Archaeological IICS system components has been created for several simple (SAMPLE, COLLECTION, CONTEXT, DISCIPLINE) and slightly more complex workflows (CONNECT, REFERTO-TOPIC, REFERTO-SPATIAL, VIEW-TO, VIEW-FROM) [5].

For the topics and content discussed here, geoscientific and archaeological information and processing are the core content. Data in this context necessarily includes applications and algorithms. Besides the above implemented features, it is optional to support any visualisation tool, processing algorithms, cartographic and mapping features and many more tools from secondary sciences, e.g., spatial algorithms and components, UDC (1-0/-9). These features can be used with the objects in any way that will be necessary to describe data and automate workflows.

C. Classification, keywords, and interfaces

The interfaces allow to use the various resources. A central element is the classification and structure of the knowledge resources as it increases the flexibility of the long-term development. Table I compares some features of classification and keywords used for object description.

Table I
UDC CLASSIFICATION AND KEYWORDS COMPARISON.

UDC	*Keywords*
Internationalisation	Methodical support, partial internationalisation
Codes	Code table support
High level of detail	Medium level of detail

Interfaces can be used in order to access and use objects. This includes filtering, combination, workflow and data processing and so on. In summation, this allows the integration of all data, objects, and resources available: scientific and discipline data, lexicographical and bibliographical data GPS data, geospatial information, processing algorithms, executable software, and many more, including realia objects. It means, we need to integrate multi-disciplinary information, allowing different views on the same context and allow even different paths for exploring knowledge.

V. LONG-TERM STRATEGY

Whereas looking from inside a traditional discipline, information seems to be complete and appears to increase slowly. On the other hand there is huge complementary information that cannot be described isolated by one discipline and tendency is increasing over the time and complexity. Table II presents the result of a reasonable categorisation that has been found from practicing the knowledge resources creation and use for several decades. It shows a more detailed compilation of categorised features and components for an expected actuality time range. In this context, the goal for long-term means > 50 years, medium-term > 15 years, and short-term < 15 years.

Table II
TIME-RANGE GOALS WITH IICS COMPONENTS (SELECTION).

Long-term	*Medium-term*	*Short-term*
Knowledge	Applications	Context
Containers	Interfaces	Sources
LX Resources	DOI, URN	URL
UDC	Converters	Media
Keywords	Active Source	Converters
Virtualisation information	Storage resources	Computing resources
Algorithms	Distributed services	Compiler, Executables
Content	Virtualisation	MPI, OpenMP
Context information	Complex implement.	Batch systems
Relations & references	Application features	Web Services
Internationalisation	OS features	Communication
Processing & workflows	Library features	Middleware

These components, described by a representative selection in Table II, can cover all aspects of knowledge creation, application, and system implementation. For example, with

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-251-6

the implementation, the resources and containers are consisting of thousands of pages. For the presentation within the following sections a small excerpt of the objects and classification can be shown. The long-term objects must be able to contain the essential knowledge, even as medium- and short-term objects cannot be preserved or made persistent as, e.g., DOI (Digital Object Identifier), URN (Uniform Resource Name), and URL (Uniform Resource Locator) will vanish and context and sources made fade away as well as OS (Operating System) features used. Therefore we have to distinct between the real instance of a DOI and URL or a context situation and a descriptive reference of these objects. These descriptive references can contain as much information and knowledge as possible (for example DOI, URL, context description, sources).

VI. Mechanisms case study

A filter chain can be used to compute resulting object sets. Based on the available system the following steps can be separated, in an example from geosciences and archaeology:

- Select topic from knowledge base (volcanology).
- Select region from results (Europe, Caribbean).
- Select volcano from results (Vesuvius, La Soufrière).
- Select object entries (geosciences and archaeology).
- Select media objects and references.
- Select application resources and interfaces.

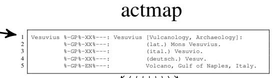
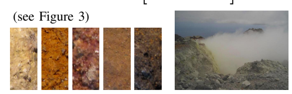
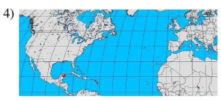
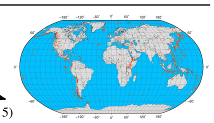
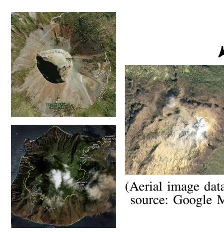
A. *Workflows and algorithms*

The knowledge resources block is the central resource in the long-term strategy. The knowledge resources can contain any kind of content. Application components can be migrated into the knowledge resources for documentation purposes and re-use. The services can access archived and historical data as well as live data and feed it into the workflows. Services interfaces allow to build complex workflows using arbitrary algorithms. The knowledge resources can be accessed from applications, which will extract suitable information and trigger the use of compute and storage resources. Objects can be selected by any algorithm, e.g., combinatory, search, and filter algorithms. The results can be delivered to a defined location or service.

B. *Resulting cross-links calculated*

The knowledge resources organisation allows to create all structures and references possible for information science purposes, besides content and taxonomy. A small subset of knowledge space objects is less suitable for providing the necessary depth for building facets and referring to appropriate multi-disciplinary objects. Therefore a basic example needs an extended knowledge environment. Table III shows an excerpt of the references and cross-links calculated from the knowledge resources for a "Vesuvius" object with some applications and originary resources, media information and illustrative examples. The secondary references (symbol: •–◦) for an object are calculated from the respective reference matrix in the knowledge resources.

Table III
EXAMPLE CROSS-LINKS WITH SECONDARY REFERENCES FOR "VESUVIUS" OBJECT AND GENERATED VIEWS (EXCERPT).

Knowledge Resources	*Applications Resources*	*Media Inform.*	*Originary Resources*
Vesuvius	LX Resources actmap	photo	samples
•–◦**volcano**		object	[on-site]
•–•–◦ **climatology**		object	–
•–•–◦ **magma chamber**	(see Listing 5)	object	[on-site]
•–•–•–◦ **Thrihnukagigur**		links	[on-site]
•–•–•–•–◦ **Þríhnúkagígur**	Transcription	links	[on-site]
•–◦**super volcano**	[interfaces]	object	[on-site]
•–•–◦ **Yellowstone**	(see Figure 3) 	object	[on-site]
•–•–•–◦ **Liparite**		object	samples
•–•–•–◦ **Rhyolithe**		object	samples
•–◦**Soufrière**		object	samples
•–◦**La Soufrière, Guadeloupe**	Movieview	video	samples
•–•–◦ **Lava sand, black**		object	samples
•–•–•–◦ **Plage de Grande-Anse**		object	samples
•–◦**Mt. Scenery, Saba**	Topicview	object	samples
•–◦**archaeology**	Photoview	photo	–
•–•–◦ **Pompeji**	(see Figure 2) 	object	museum
•–•–◦ **Herculaneum**		object	museum
•–•–◦ **Hephaistos**		object	–
•–•–◦ **Vulcanus**	(see Figure 4) 	object	–
•–•–•–◦ **Puteoli**		object	–
•–•–◦ **Zipacná**		object	–
•–•–◦ **Cabrácan**		object	–
. . . [cut] . . .	Objectmap		–
•–◦**Volcano compendium**	Volcanomap 		
•–•–◦ **VNUM**			
•–•–◦ **Location**			
•–◦**Archaeological compendium**			
•–◦**Geological table**	(see Figure 5)		
•–◦**Stone table**	Stonetab	object	samples
•–◦**Mineral table**	Mintab	object	samples
•–•–◦ **Periodic table of elements**	Elemview	object	samples
. . .  (Aerial image data, e.g., source: Google Maps)	Mapview	image	–
	Satelliteview	service	–
			–
	Index	service	–
	Dynasearch	service	–
	Dynacompute	service	–
	Dynaprocessing	service	–
	Dynaindex	service	–
	Dynatypeset	service	–
	Dynastat	service	–
[compute and storage resources]	[workstation]		

The cross-links are extracted from the long-term knowledge resources and are usable with available data by any application components. Available example instances from the application components set, media resources, and originary sources (realia) are listed. In this context cross-linked subentries, keys and, for example, in case of volcanic features the classification (UDC) and the Volcano Number (VNUM) can be integrated in the analysis for computation and processing via the reference to the appropriate container, the volcano compendium. Applications and interfaces (e.g., Active Source, actmap [4]) can access any available data, as creating index compilations, retrieving online maps and satellite imagery. Media are, e.g., video data and photo images. Originary sources of objects are, e.g., realia, in the case with the volcanic material here, the objects are volcanic samples, respective stones. These objects are referred to in an appropriate archive, e.g., a collection, library or museum. The bottom row shows how the services and resources are used (e.g., local workstation access, special compute and storage resources).

C. *Resulting object classifications and processing*

Listing 1 shows a simple UDC-Context sample subentry from the LX Resources. This example expresses a basic “Europe : America” relation.

```
UDC-Context:(4):(7)
```

Listing 1. UDC-Context sample (LX Resources).

Classification patterns are suitable to select objects with any algorithm and pattern matching rules. Listing 2 shows a basic grep-filter within a classification set of UDC samples.

```
... | egrep "\([0-9]\):\([0-9]\)" | grep "(4):(7)"
```

Listing 2. Search grep sample within classification (LX Resources).

The result is a set of objects with a context relation between the continents Europe and America. Listing 3 shows an excerpt of an UDC table used with the examples.

```
%%IML: UDC55     :: Earth Sciences. Geological sciences
%%IML: UDC56     :: Palaeontology
%%IML: UDC57     :: Biological sciences in general
%%IML: UDC911.2  :: Physical geography
%%IML: UDC902    :: Archaeology
%%IML: UDC903    :: Prehistory. Prehistoric remains, artefacts, antiquities
%%IML: UDC904    :: Cultural remains of historical times
%%IML: UDC930.85 :: History of civilization. Cultural history
%%IML: UDC"63"   :: Archaeological, prehistoric, protohistoric periods, ages
%%IML: UDC(4)    :: Europe
%%IML: UDC(7)    :: North and Central America
%%IML: UDC(23)   :: Above sea level. Surface relief. Above ground generally.
%%IML: UDC(24)   :: Below sea level. Underground. Subterranean
%%IML: UDC=12    :: Italic languages
%%IML: UDC=14    :: Greek (Hellenic)
%%IML: UDC=84/=88:: Central and South American indigenous languages
```

Listing 3. UDC classification table, English (LX Resources, excerpt).

For a Topicview, corresponding object classes are processed and media objects are computed. Listing 4 shows the core data generated from the filtered media data, processing images for a sample Topicview as listed in Table III.

```
create_archaeology_planet_view_topic.sh \
   volcano_guadeloupe_soufriere_viewto.jpg \
   volcano_guadeloupe_soufriere_viewfrom.jpg \
   volcano_saba_mtscenery_viewto.jpg \
   volcano_saba_mtscenery_viewfrom.jpg
```

Listing 4. Generated core data for Topicview processing.

As the knowledge resources’ objects carry references to any kind of detailed processable data, distribution maps and satellite views can be computed and passed on.

D. *Resulting object entries on geosciences and archaeology*

The following object entries are excerpts from the calculated cross-links table (Table III). The excerpts contain some, structure, UDC classification, keywords, references, and satellite image reference. The references for the geopositioning are created via classification. They can be used for any purpose where geopositioning will seem interesting. Listing 5 shows an excerpt of an LX Resources object entry [1], “Vesuvius” volcano.

```
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\UD{m}.
         The volcanic activity in the region is observed by the Oservatorio
         Vesuviano. The Vesuvius area has been declared a national park on
         \isodate{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
         %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(4+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
```

Listing 5. Knowledge resources – object entry “Vesuvius” volcano.

The example contains a reference and VNUM for the Vesuvius volcano, various secondary objects, UDC classification, satellite image reference (Satelliteview in Table III). It refers to “Soufriere”, “La Soufrière”, and “Mt. Scenery”. Listing 6 shows an excerpt of the “Soufriere” object entry.

```
Soufriere    [Vulcanology, Geology]:
             A common name for a volcanic feature resulting from the
             french term for \periref{tgt:PeriSulfur}{Sulfur}.
             The name soufriere is used for a volcanic crater
             or other area in combination with solfataric activity.
             The name is mostly used in French speaking regions,
             especially in the West Indies.
             Very well known are, for example:
             La Soufrière volcano, Guadeloupe, F.W.I.
             Soufriere Hills, F.W.I.
             Soufriere St. Vincent, F.W.I.
             Syn.: Soufrière
             s. also La Soufrière, F.W.I., volcano, seismology
             %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(7+23)=84/=88
```

Listing 6. Knowledge resources – secondary object entry “Soufriere”.

This secondary object entry, “Soufriere”, also refers to the La Soufrière volcano (Listing 7), which itself refers to various data and objects, e.g., satellite image references.

```
La Soufrière [Vulcanology, Geology]:
             La Soufrière volcano, Guadeloupe, F.W.I.
             Volcano Type: Stratovolcano,
             Country: France,
             Subregion Name: West Indies, Caribbean,
             VNUM: 1600-06=,
             Summit Elevation: 1467\UD{m}.
             Syn.: Soufriere
             s. volcano
             s. also Soufriere, F.W.I., lava, lava sand, OVSG
             %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(7+23+24)=84/=88
             %%IML: GoogleMapsLocation: http://maps.google.com/?ie=UTF8&ll
              =16.043153,-61.663374&spn=0.003088,0.003262&t=k&z=18&vpsrc=6&
              lci=weather
```

Listing 7. Knowledge resources – secondary object “La Soufrière”.

The secondary object entry “Mt. Scenery” (Listing 8) also contains classifications and media, and further data references for the Mt. Scenery volcano on Saba. Extracted examples are volcano type, VNUM, region, status, elevation, and UDC classification views as well as the geo-references, which with this request are used to automatically compute views and distribution maps for classified objects in the result matrix. The classification groups themselves show references to associated objects. The data and media object in a processed reference chain can be used for further

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-251-6

analysis, creating special features. That way, using UDC classifications, e.g., places from a region or context that can be associated with volcanology and associated with archaeological sites can be selected and media objects can be processed and realia referred.

```
Mt. Scenery  [Vulcanology, Geology]:
             Volcano, Saba, Netherlands Antilles, D.W.I.
             Volcano Type: Stratovolcano,
             Country: Netherlands,
             Subregion Name: West Indies, Caribbean,
             VNUM: 1600-01=,
             Volcano Status: Historical,
             Last Known Eruption: in or before \isodate{1640}{}{},
             Summit Elevation: 887\UD{m}.
             %%IML: UDC:[55+56+911.2]:[902+903+904]:[57+930.85]"63"(7+23+24)
              =84/=88
             %%IML: GoogleMapsLocation: http://maps.google.com/maps?f=q&
              source=s_q&hl=en&geocode=&q=mt+scenery,+saba+netherlands+
              antilles,+google+maps&aq=&sll=17.633225,-63.236961&sspn
              =0.048997,0.052185&vpsrc=0&t=h&ie=UTF8&hq=mt+scenery,+saba+
              netherlands+antilles,&hnear=&z=14&lci=weather
             s. also Saba, D.W.I., volcano, seismology
```

Listing 8. Knowledge resources – object entry "Mt. Scenery".

Dynamical components can even benefit from precalculation and precomputation of objects. This includes precalculated classification and weights ("PreUDC"). The following section presents examples calculated from the above classified objects and the figures are showing the results of selected attributes, including the classification and geo-references used basic visualisation.

E. Resulting features selection for cross-links processing

The following (Figure 2) is an excerpt of the secondary objects computed above for the Caribbean region volcanoes and a selection with UDC "`(23)`", "`(24)`".

Figure 2. Topicview – volcanoes, La Soufrière (left), Mt. Scenery (right), VIEW-TO (green), VIEW-FROM (blue).

Figure 3 illustrates the computed objects (Topicview), e.g., here volcanic samples after processing, all showing the variety of material from the top of the La Soufrière volcano.

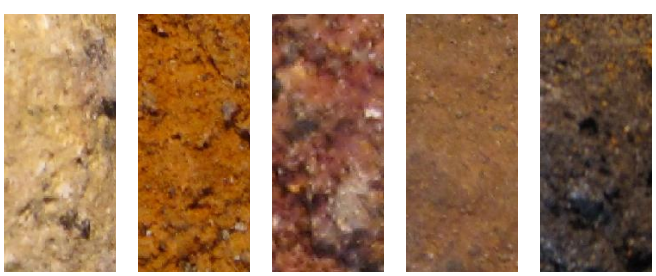


Figure 3. Topicview – related volcanic samples (La Soufrière, 2011).

Any of these objects being part of the resulting matrix for a request, e.g., photos for object entries as well as media data for physically available samples, have been found via references and UDC from the knowledge base. The realia references for the objects refer to a collection where the samples are stored. Further analysis for the samples is available via the knowledge resources. Figure 4 shows the geolocations [1] for computed geoscientific and archaeological object samples on a configurable object map.

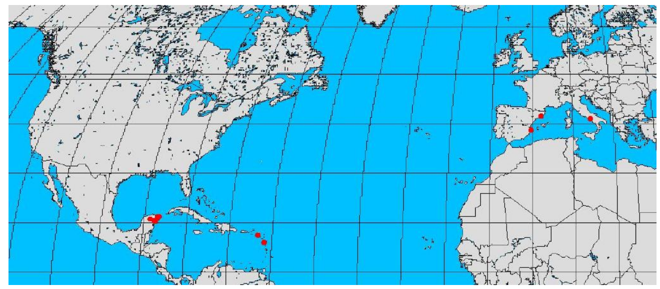


Figure 4. Objectmap – computed map for related objects (red, excerpt).

A sample distribution of volcanic features is depicted in Figure 5. It shows a comparison of volcanic data in a projection identical to the computed Objectmap (Figure 4). Although the knowledge matrix of this example is most complex (Table III), the workflow for producing a view can be specified very easy like for spatial presentation.

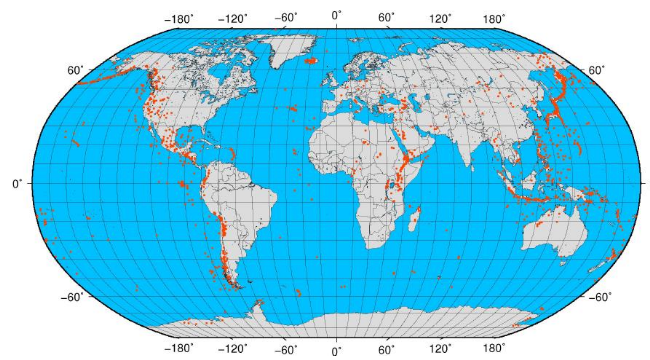


Figure 5. Volcanomap – worldmap of referenced volcanoes (orange).

The map generated with the workflow as described with the case study presents the related objects from the context available in the geophysical research database. It can visualise various aspects of the classified objects. In this case of volcanoes and geological samples a reasonable view is the spatial distribution of the referenced selection.

Anyhow it must be emphasised that the number of possible views is not limited, neither from the knowledge base nor from the implementation. Spatial and cartographic methods provide only a very restricted tool set for supporting sciences for their complex tasks. For example, more complex examples from the same context could use more advanced presentation methods than available from spatial procedures. As it is obvious from this, the implementation of the knowledge resources architecture can be used for any purpose.

With the suggested workflow, the objects from the knowledge resources can be processed by any means like phonetic search, e.g., via classical or modified Soundex algorithms.

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-251-6

This includes the flexible development of a non-limited number of extensions for dynamical search and analysis. It, too, provides a multiplicity of granularity regarding objects and classification.

Any features and data shown, based on the knowledge resources [1] and further sources, even if much less structured like online encyclopedia material, are resulting from the request and workflow, e.g., selection of classification, topics, object, secondary data, area, map projection, applications and so on [9], [10], [11].

One possible example of an algorithm for the interface workflow, with one request iteration is: Knowledge base request, keyword filtering, object processing, UDC filtering, object element processing, object container retrieval, media retrieval, media and container processing, building resulting media, visualisation, provisioning results. This can be used to create multi-disciplinary text and media results, e.g., dynamical distribution maps, from requests, using calculation, processing, and computation of objects. A workflow can enter the knowledge matrix from different directions, e.g., from topic to related topic or from overview to detailed view as well as vice versa.

VII. Evaluation

The integration of structure and classification allows to use the benefits of algorithms like filtering for any possible use of processing and computing. Structuring the content and context documentation allows a flexible balance for redundancy of data and compute requirements for various application scenarios, even with identical data.

The facetted classification and multi-disciplinary data have proved to provide significant benefits for knowledge reuse and discovery. This includes various ways of describing aspects correctly. From one view a glass of water is half full. From another view the same glass of water is half empty. The two groups representing the classical views might argue that the other view is unintelligible. Both are generally not good as they only represent views. An alternative view will be describing the status giving a filling percentage. In addition this reduces the limitation of unprecise references.

Most content, tasks, and developments handled with information and computing systems are not suitable for any long-term use. The use of the universal knowledge resources and collaboration framework has shown to be very flexible and extendable with implementations and technologies over several decades.

It has been found that standard search and pattern recognition on information is by far not sufficient to gain reasonable results for long-term knowledge herding and evaluation processes. In contrast, the implementation shows excellent results with opening multi-discipline data for IICS, advanced computing, and processing. Statistically, filtering 1 GB of unstructured data delivers less quality than using 10 MB structured classified knowledge base data. The Quality of Data (QoD) must be drastically improved in order to get better results. This can help to reduce compute times, storage volume, and besides overall costs it can help to decrease energy consumption in the end. Using UDC in this context, the availability of a full UDC catalog, and an implementation allowing classification views, combined classification, and ranking priority has proven to drastically increase the QoD. With multi-disciplinary networks, there is even need for a tolerance of individual classification.

In common environments it is only feasible to do one implementation for a specific application, as has been done with these components. Anyhow it has been possible to implement the applications on various architectures providing different resources. Workflows support the use of remote resources (Table IV). In case of a 1000 knowledge-objects reference chain, with 1–10 elements per object, performance will increase much with low latencies.

Table IV
Workflow processes (remote, Ethernet, 1000 nodes).

Remote Workflow Process	*Elements*	*Response Time*
Knowledge base request	1000	5 s
Processing (object, media)	10	7 s
Building result	10	5 s
Visualisation	2	25 s

When using one of the described very basic application scenarios on a certain resources architecture the efficiency mostly depends on the decision for the depth of the cross-links to be considered and on the processing requirements for the media data for the originary resources. With current sizes for digital photos and a low depth of five to ten for the cross-links a medium sized application can easily use about one-hundred parallel processes. On a common compute resource without a queue configured for the jobs the response time will be less than a minute.

That way, implementing components with IICS on many compute nodes can profit from using various technologies as suited for different purposes, using task and thread parallelism to the extend needed to handle a problem remotely: *High level*: Integrated Systems, collaboration frameworks, Partitioned Global Address Space (PGAS) models. *Virtualisation level*: Parallel Virtual Machine (PVM). *Low level*: Message Passing Interface (MPI) and OpenMP.

VIII. Conclusion and future work

It has been demonstrated how complex systems for multi-disciplinary documentation and computing can be built, developed, and extended based on creating long-term-knowledge resources supported by a universal classification and implementing IICS systems. This paper presented the successful implementation of a new universal framework for an integrated system, integrating knowledge resources and implementation components for long-term knowledge creation and use, including the facilities for High End Computing and processing resources.

The geoscientific and archaeological knowledge resources have been, structured, extended, and developed for several decades now, having been successfully used with various technology over time. Huge benefits creating new instances of objects and components result from enabling a long-term stepwise development for all parts of the knowledge and application space and a free extendability of the knowledge base. The previous work this implementation is built on has been discussed.

The architecture allows any kind of documentation and algorithm for content, context, information and resources usage. The services and resources usage is very economic and only limited by the limiting implementation factors, e.g., capacities and policies. This solution goes far beyond data and text mining or image analysis and pattern recognition. As shown, classification, as well as spatial data should be integrated with the objects. In no case is it suitable regarding the long-term goals of knowledge creation to "fix" knowledge objects with an application or implementation, neither simple or complex, nor closed or open licensed.

The comparison showed that the possibility of combining methods (UDC, keyword, full-text analysis) does lead to unique benefits. Comparable precision, reliability, performance, and scalability is not available from any isolated method. For any advanced knowledge resources and improved QoD, a flexible classification is undispensable. Bringing the integration of universal classification and IICS into wider acceptance can provide a time-capsule against the transience of knowledge and open new synergetic long-term possibilities.

Complex systems can be created and extended over the necessarily long periods of time, using IICS and UDC. Advanced scientific computing is supported by interfaces, accessing compute and storage resources. Further it has been shown what information and knowledge on content and context can be preserved for medium- and long-term usage even for large complexity of an overall system.

The basic architecture has been presented using a long-term knowledge base (LX), documentation, and classification of objects, the "Collaboration house" framework, flexible algorithms, workflows and dynamical and Active Source components for creating future IICS. Besides that, there is a strong demand for future education and teaching in all disciplines of academia and research in order to mediate and disseminate the basics of knowledge creation and classification.

ACKNOWLEDGEMENTS

We are grateful to all national and international partners in the GEXI cooperations for the innovative constructive work and to the scientific colleagues at the Leibniz Universität Hannover, the Institute for Legal Informatics (IRI), and WWU Münster, sharing experiences on ZIV, HLRN, Grid, and Cloud resources and for participating in fruitful case studies as well as the participants of the EULISP Programme for prolific scientific discussion over the last years. We do thank the international colleagues from geosciences, informatics, and archaeology in the present collaborations for constructive feedback and proof-reading this paper.

REFERENCES

[1] LX-Project, "LX Foundation Scientific Resources," 2012, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX (Information) [accessed: 2012-09-29].

[2] C.-P. Rückemann, "Integrating Information Systems and Scientific Computing," *Int. Journal on Advances in Systems and Measurements*, 2012, ISSN: 1942-261x, (to appear).

[3] "Geo Exploration and Information (GEXI)," 1996, 1999, 2010, 2012, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/index.html#GEXI (Information) [accessed: 2012-09-29].

[4] C.-P. Rückemann, "Beitrag zur Realisierung portabler Komponenten für Geoinformationssysteme. Ein Konzept zur ereignisgesteuerten und dynamischen Visualisierung und Aufbereitung geowissenschaftlicher Daten," Diss., WWU, Münster, Deutschland, 2001, 161 (xxii+139) S., URL: http://www.user.uni-hannover.de/cpr/x/publ/2001/dissertation/wwwmath.uni-muenster.de/cs/u/ruckema/x/dis/download/dis3acro.pdf [accessed: 2012-09-11].

[5] C.-P. Rückemann, "Enabling Dynamical Use of Integrated Systems and Scientific Supercomputing Resources for Archaeological Information Systems," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2012), October 21–26, 2012, Venice, Italy*. XPS, Xpert Publishing Services, 2012, pp. 36–41, Rückemann, C.-P., Dini, P., Hommel, W., Pankowska, M., and Schubert, L. (eds.), ISBN-13: 978-1-61208-226-4, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2012_3_10_10012 [accessed: 2012-11-18].

[6] Universal Decimal Classification Consortium (UDCC), "Universal Decimal Classification (UDC)," 2012, URL: http://www.udcc.org [accessed: 2012-02-19].

[7] "Universal Decimal Classification (UDC)," 2012, Wikipedia, URL: http://en.wikipedia.org/wiki/Universal_Decimal_Classification [accessed: 2012-09-09].

[8] B. A. Worley, "Knowledge Discovery from Data," *International Panel on Future High End Systems: Chances and Challenges for Intelligent Applications and Infrastructures, October 22, 2012, The International Conference on Advanced Communications and Computation (INFOCOMP 2012), October 21–26, 2012, Venice, Italy*, 2012, URL: http://www.iaria.org/conferences2012/ProgramINFOCOMP12.html [accessed: 2012-11-17].

[9] B. Steinberger, "Plumes in a convecting mantle: Models and observations for individual hotspots," *JGR*, vol. 105, pp. 11 127–11 152, 2000.

[10] "GMT - Generic Mapping Tools," 2012, URL: http://imina.soest.hawaii.edu/gmt [accessed: 2012-09-02].

[11] GDAL Development Team, *GDAL - Geospatial Data Abstraction Library*, Open Source Geospatial Foundation, 2012, URL: http://www.gdal.org [accessed: 2012-07-15].