International Journal on Advances in Systems and Measurements, vol 6 no 1 & 2, year 2013, http://www.iariajournals.org/systems_and_measurements/

Archaeological and Geoscientific Objects used with Integrated Systems and Scientific Supercomputing Resources

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 methods and results from combining Integrated Information and Computing System components with classification for the purpose of enabling multi-disciplinary and dynamical use of information systems and supercomputing resources for Archaeological Information Systems. Focus is on soft criteria, structures, and classification for knowledge discovery for sustainable, long-term knowledge resources. The essential base are a flexible collaboration framework, suitable long-term documentation, structuring and classification of objects, computational algorithms, object representations, and workflows as well as portable application components like Active Source. Case studies of the successful implementation of integration of archaeology and geosciences information and facilitation for dynamical use of High End Computing resources are discussed. The implementation shows how the goal of integrating information and systems resources and advanced scientific computing for multi-disciplinary applications from natural sciences and humanities can be achieved by creating and using long-term knowledge resources.**

Keywords–Integrated Systems; Scientific Supercomputing; Knowledge Resources; Archaeology; Geosciences; Information Systems; Phonetic Algorithms; High Performance Computing.

I. INTRODUCTION

The target of this development is sustainable long-term knowledge resources providing information found by necessarily sophisticated workflows considering content and context. This has to go along with systematically structuring system components and information and describing content and context of objects. With archaeology the objects are commonly handled in a data collection different from the natural sciences objects. The collection and description normally shares no geoscientific, physical, and secondary data, e.g., from natural sciences.

Technology and components are used for digital library components, classification of objects, and realia. Nevertheless, it is important for many use cases in geosciences and archaeology to enable a dynamical use of Integrated Systems and computing resources [1]. In order to overcome many of the complex scientific impediments in prominent disciplines we do need mighty information systems but the more they are used for interactive use they show up needing capabilities for the state-of-the-art in dynamical computing. The studies and implementations of Integrated Information and Computing Systems (IICS) have shown a number of queuing aspects and challenges [2], [3]. In the case if archaeological information systems needed for multi-disciplinary investigation the motivation is the huge potential of integrative benefits and even more pressing that archives are needed for multi-disciplinary records of prehistorical and historical sites while context is often being changed or destroyed by time and development. Besides the academic, industrial, and business application scenarios in focus of the Geo Exploration and Information collaborations (GEXI) [4] in order to integrate the necessary computing facilities with these systems, on the technical side the recent implementations for spatial control problems, e.g., for wildfire control [5], integrating GIS, and parallel computing are promising candidates for future support. This research paper especially contributes to the most important aspect of soft criteria in creating knowledge resources and implementing effective knowledge discovery.

This paper is organised as follows. Sections II and III introduce the basic knowledge resources and the necessary long-term investments. Section IV shows the essential prerequisites of information and structure for the information and computing systems. Sections V and VI describe the results from the development of “silken criteria” and presents examples from phonetic support. Sections VII and VIII show a workflow from these developments and explain the importance of these criteria for the context. Section IX presents the high-level results for the computation and parallelisation with these workflows. Sections X to XIV describe and evaluate the resulting implementation for an Archaeological Integrated Information and Computing Systems and computation results from the components, based on the knowledge resources and digital library examples. Section XV summarises the conclusions and future work.

II. KNOWLEDGE RESOURCES

Knowledge resources provide the universal base for using information and computing resources for a multitude of purposes. They contain systematically gathered, structured and documented content and context on any kind of information, object, sources, and tools.

This includes systematically structuring system components and information and describing content and context of objects. The architecture and structure enables to use any kind of workflow, e.g., filter stacks using flexible algorithms on different type of content and context. Information and data can be data-mined, analysed, retrieved, and used, e.g., for processing, computing or typesetting by sophisticated workflows considering any qualities or properties of the material. Examples for the material are data sets from natural sciences, documentation texts on multi-disciplinary topics, descriptive texts on humanities, media data, photo documentation on objects, e.g., from digital libraries, and visualised data.

III. Sustainable long-term investments

Although there is some overlap, investments can be categorized in investments for disciplines, services, and resources. For long-term scientific goals, the most significant investment is in knowledge resources. As disciplines have to care for their content and results these may be the investments being closest to the work within disciplines.

Nevertheless, there will be a number of components, e.g., algorithms and applications, which will directly be cared for by disciplines. Services are regulary provided by specialised groups. Computing and storage resources can be provided by various groups, as long as the necessary size and performance is not at the top edge.

All the developments presented in this paper can be considered to be tightly coupled to the knowledge resources, therefore being of close interest for participating disciplines: Silken criteria, parallelisation, workflow and context, information structure, classification, integrated information and computing systems. The investments in the knowledge resources have proved to provide highest sustainability for over twenty-five years now.

IV. Information and structure

It must be emphasised that the complexity of the ecosystem of algorithms and disciplines necessary to achieve an integration of multi-disciplinary information and components is by nature very high so besides the system components we have not only to integrate unstructured but highly structured data with a very complex information structure.

The overall information is widely distributed and it is sometimes very difficult and a long lasting challenge even to get access to a few suitable information sources. The goal for these ambitions is an integrated knowledge base for archaeological geophysics. Example data resources and methods are [6], [7], [8], [9], [10], [11], [12]. For all components presented, the main information, data, and algorithms are provided by the LX Foundation Scientific Resources [13].

Structuring information requires a hierarchical, multilingual and already widely established classification implementing faceted analysis with enumerative scheme features, allowing to build new classes by using relations and grouping. This is synonym to the Universal Decimal Classification (UDC) [14]. In multi-disciplinary object context a faceted classification does provide advantages over enumerative concepts. Composition/decomposition and search strategies do benefit from faceted analysis. It is comprehensive, and flexible extendable. A classification like UDC is necessarily complex but it has proved to be the only means being able to cope with classifying and referring to any kind of object.

V. Silken criteria: Phonetic support

Common means of knowledge exploitation provide string search, precise mathematical algorithms for selections and so on. These are rather sharp with their precision. Even string searches based on regular expressions using advanced means of wildcards are limited in terms of not simply matching the characters but the meaning or context.

For increasing the quality of exploiting knowledge resources we have to build sophisticated means of searching and filtering information objects. For example, if knowledge resources contain more features, these can be used in combination:

- Structure,
- Classification,
- Language distinction,
- Pattern recognition,
- "Sound" recognition, ...

The entirety of the knowledge resources being part of the LX Foundation Scientific Resources [13] does provide unique means of collective use of features that can be used for knowledge based recognition.

Building applications based on the integrated features, this results in synergy on the one hand and in a much higher Quality of Data (QoD) on the other hand. For example, with search requests, the percentage of information used with the resulting matrix is much higher with integrated features. Standard search includes about 20–50 percent of the available first level information. Integrated search allows to gain up to over 90 percent of suggested information in the first level and about the same for the second and following levels. This does require much higher demands for computation, with most applications even in interactive time range, but this should not be a problem today.

In many of the applications built on knowledge resources, an uncertainty for various attributes is necessary. Algorithms solely being precise as well as those implementing an uncertainty have shown drawbacks when being used for increasing the quality of the results. The solution for many application is to implement sequences of those types of algorithms.

So, with the above mentioned features of the knowledge resources, the individual strengths are that, for example:

- Structure can integrate scientific names, e.g., botanical names, with commonly used names whatever they might be matching.
- Pattern recognition can be used to find matching objects on string basis, e.g., from exactly matching character strings.
- Classification can help to find context as well as choosing object or filtering besides any pattern or structure matching, e.g., with UDC codes.
- Language distinction can be used for supporting classification and pattern matching as well as typesetting and hyphenation support, mostly by improving the precision of meaning and context, e.g., generating publishing objects.
- Sound recognition can help find homophones and comparable objects, e.g., searching and selecting additional paths of knowledge discovery to follow in a workflow or filter process.

So, with these resources, even elementary modules for sound and pattern recognition can be of huge benefits when being integrated with the other methods.

VI. Supporting Silken Selection

The knowledge resources can be used by any algorithm suitable for a defined workflow. One of the available module implementing a silken selection based on the Soundex principle is the `knowledge_sndx_standard` application. The historical Soundex [15] is a phonetic algorithm for indexing names by sound. The goal with this algorithm is to encode homophones so that they can be represented by the same resulting code in order to match persons' name despite differences in writing and spelling [16]. The basic algorithm mainly encodes consonants. Vowels are not encoded unless being a first letter. The U.S. Government is using a modified modern rule set [17] for purposes in census and archives. The original intention was to catch the English pronunciation, anyhow there are many different implementations in use today.

Listing 1 shows a Perl source code used in the `knowledge_sndx_standard` module, modelled after the standard Perl implementation [18], for computing LX Soundex codes [19], being available based on different programming concepts [20], [21]. The various workflows can define and integrate their own Soundex codes for different purposes and topics.

```
#!/usr/bin/perl
#
# knowledge_sndx_standard -- (c) LX Project -- CPR 1992, 2012
#

$string=$ARGV[0];
$sndx_nocode = undef;

sub knowledge_sndx_standard
{
  local (@s, $f, $fc, $_) = @_;
  push @s, '' unless @s;

  foreach (@s)
  {
    $_ = uc $_;
    tr/A-Z//cd;

    if ($_ eq '')
    {
      $_ = $sndx_nocode;
    }
    else
    {
      ($f) = /^(.)/;
      tr/AEHIOUWYBFPVCGJKQSXZDTLMNR/00000000111122222222334556/;
      ($fc) = /^(.)/;
      s/^$fc+//;
      tr///cs;
      tr/0//d;
      $_ = $f . $_ . '000';
      s/^(.{4}).*/$1/;
    }
  }

  wantarray ? @s : shift @s;
}

$code = knowledge_sndx_standard $string;
print ("SNDX-standard:$code:$string\n");

##EOF:
```

Listing 1. LX Soundex SNDX-standard module Perl source code.

The next examples are multi-disciplinary objects from one context, linked by the references in the knowledge resources. If the `SNDX-standard:` prefix is left out in the following examples, the code refers to this standard code.

A. Geology and volcanology

Listing 2 shows some computed LX Soundex codes for the La Soufrière volcano and the reference-internal comparable sound occurrences. The code unifies a number of different versions primarily linked by the prefix but classified by the object classification.

```
L216:La_Soufriere
L216:La_Soufri{`e}re
L216:La_Soufrière
```

Listing 2. SNDX-standard codes for La Soufrière.

The same is true for the following related object. Listing 3 shows computed LX Soundex codes for the Vesuvius and the reference-internal comparable sound occurrences.

```
V210:Vesuv
V210:Vesuvio
V212:Vesuvius
```

Listing 3. SNDX-standard codes for Vesuvius volcano.

Both of these object examples are referring to the volcano database.

B. Archaeology

The archaeological objects, too, very well fit with these algorithms. This is true for a large number of more than 95 percent of classified entries. Listing 4 shows some computed LX Soundex codes for Yucatán and the reference-internal comparable sound occurrences.

```
Y235:Yucatan
Y235:Yucat'an
Y235:Yucatán
```

Listing 4. SNDX-standard codes for Yucatán.

Listing 5 shows a number of computed LX Soundex codes for Chichén Itzá and the reference-internal comparable sound occurrences.

```
C250:Chichén
C253:Chich'en_Itz'a
C253:Chichen_Itza
C253:Chichén_Itzá
```

Listing 5. SNDX-standard codes for Chichén Itzá.

Listing 6 shows computed LX Soundex codes for Cobá and the reference-internal comparable sound occurrences.

```
C100:Coba
C100:Cob'a
C100:Cobá
```

Listing 6. SNDX-standard codes for Cobá.

C. *Biology and botanics*

For any of the objects there may be different spellings or even different terms. This means that there are, e.g., botanical names, which are not homophonetically near to the other terms. Listing 7 shows some computed LX Soundex codes for the Chiricote and the reference-internal comparable sound occurrences.

```
G260:Geiger
C623:Chiricote
C623:Ciricote
Z623:Ziricote
C630:Cordia
```

Listing 7. SNDX-standard codes for Chiricote.

The higher variability of codes from the knowledge resources is a good source for calculating new trees for the knowledge discovery workflow.

D. *Names and sources*

Searching the knowledge resources for "geology, volcanology, and earthquake" delivers a person "Leibniz" in the result Matrix, referring to one of the early statements that volcano activity can result in earthquakes. As the Leibniz object carries a large number of pseudonyms, it can be interesting to follow these as non-explicit references.

An algorithm supports building groups of pseudonyms. Listing 8 shows a computed LX Soundex code for a selection of names used in context with Gottfried Wilhelm Leibniz (1646–1716) and their reference-internal comparable sound occurrences, as computed for the result matrix.

```
SNDX-standard:C260:Caesar
SNDX-standard:C262:Caesarius
SNDX-standard:C265:Caesarinus
SNDX-standard:F612:Freybach
SNDX-standard:F623:Fuerstenerius
SNDX-standard:F623:Fürsteneer
SNDX-standard:F623:Furstenerius
SNDX-standard:F623:Fürstenerius
SNDX-standard:G163:Goffredo
SNDX-standard:G244:Guglielmo
SNDX-standard:G316:Godefridus
SNDX-standard:G316:Godefroy-Guillaume
SNDX-standard:G316:Godfridus
SNDX-standard:G316:Godofredus
SNDX-standard:G316:Godofridus
SNDX-standard:G316:Gotfrid
SNDX-standard:G316:Gotfrids
SNDX-standard:G316:Gothofredus
SNDX-standard:G316:Gotofredus
SNDX-standard:G316:Gottefridus
SNDX-standard:G316:Gottfredus
SNDX-standard:G316:Gottfrid
SNDX-standard:G316:Gottfried
SNDX-standard:G316:Gottofredus
SNDX-standard:G426:Gallo-Graecus
SNDX-standard:G445:Guilelmus
SNDX-standard:G445:Guilielmus
SNDX-standard:G445:Guillielmus
SNDX-standard:G445:Gulielmus
SNDX-standard:G620:Georg
SNDX-standard:G622:Georgius
SNDX-standard:G622:Graecus
SNDX-standard:G655:Germano
SNDX-standard:G655:Germanus
SNDX-standard:J235:Justiniano
SNDX-standard:L152:Leibnics
SNDX-standard:L152:Leibniz
SNDX-standard:L152:Leibnizius
SNDX-standard:L152:Leibnüz
SNDX-standard:L152:Leibnuzius
SNDX-standard:L152:Leibnüzius
SNDX-standard:L153:Laipunitsu
SNDX-standard:L153:Leibnitio
SNDX-standard:L153:Leibnitius
SNDX-standard:L153:Leibnits
SNDX-standard:L153:Leibnitz
SNDX-standard:L153:Leibnitzius
SNDX-standard:L153:Leibnütz
SNDX-standard:L215:Lajbnic
SNDX-standard:L215:Lejbnic
SNDX-standard:L315:Lithvanus
SNDX-standard:L352:Lithuanus
SNDX-standard:R114:Republicanus
SNDX-standard:R153:Raibunittsu
SNDX-standard:S125:Sibisimilis
SNDX-standard:S516:Semper
SNDX-standard:U421:Ulicovius
SNDX-standard:V445:Vilhelm
SNDX-standard:V632:Veridicus
SNDX-standard:W445:Wilhelm
```

Listing 8. LX SNDX-standard codes for "Leibniz" pseudonym parts.

The result shows that the name-Soundex algorithm delivers several phonetical groups. Distinction criteria for modelling the results can be based on considering knowledge resources' structure, attributes, and features, e.g., language, topic context, and name-string order.

Here, the most frequent groups are G316, G244, G163, G445, W445, L152, L153, L215. On the one hand, these obviously correspond with different spellings of the real name. On the other hand, pseudonym name parts are especially carrying codes as C260, C262, C265, F612, F623, G622, G426, G655, J235, L315, L352, R114, R153, S125, S516, U421, V632. Further, if necessary for a workflow, it is as well possible to handle phonetical variances and pseudonym names separately, oeven with separate phonetical algorithms.

Listing 9 shows some essential modifications for the SNDX-latin module `knowledge_sndx_latin` compared to the SNDX-standard (Listing 1), to be used with these groups of objects.

```
      tr/AEHIOUWYBFPVCGJKQSXZDTLMNR
       /00000000111122022222324556/;
```

Listing 9. LX Soundex SNDX-latin modification for SNDX-standard.

Listing 10 shows a computed LX Soundex code for an excerpt selection as above but with the SNDX-latin module.

```
SNDX-latin:L152:Laipunitsu
SNDX-latin:L152:Lajbnic
SNDX-latin:L152:Leibnics
SNDX-latin:L152:Leibnitio
SNDX-latin:L152:Leibnitius
SNDX-latin:L152:Leibnits
SNDX-latin:L152:Leibnitz
SNDX-latin:L152:Leibnitzius
SNDX-latin:L152:Leibniz
SNDX-latin:L152:Leibnizius
SNDX-latin:L152:Leibnütz
SNDX-latin:L152:Leibnüz
SNDX-latin:L152:Leibnuzius
SNDX-latin:L152:Leibnüzius
SNDX-latin:L152:Lejbnic
```

Listing 10. LX SNDX-latin codes for "Leibniz" pseudonym name parts (excerpt) showing the harmonised codes.

The newly created algorithm has harmonised the codes L152, L153, L215 for the "Leibniz"-object regarding 'z' and 't' as well as 'i' and 'j' to become SNDX-latin:L152. In order to benefit from the improvements with algorithms, objects can carry any references to these algorithms. For the disciplines creating the content and references it is important not only to see the result matrix but also the reasons for the codes and ranking and to be able the modify the source codes with any objects.

VII. Workflow and Silken Criteria

The workflow for applying these algorithms for an enriched result matrix is as follows:

1) Object search using string and classification criteria on the knowledge resources and references results in primary result matrix.
2) Object search using smooth, silken criteria, e.g., Soundex, on attribute- selected content in the primary result matrix results in secondary result matrix.
3) References to object from the secondary result matrix are used to search objects from the knowledge base and references in order to create a tertiary result matrix.
4) The tertiary result matrix is integrated with objects from all steps and a defined ranking is used to create the final result matrix.

Methods include the structure of objects, language attribute, transliterations, transcriptions, synonyms, references and so on. In most cases these features are precisely defined. The silken support is provided by an algorithm defined for and by the user application within the scenario. This algorithm, by concept, is designed to enable a use case specific implementation.

VIII. Workflow and Context

In the regular expression - knowledge resources workflow (workflow 1), the result will be based on the chain "*Volcano - Vesuvius - Vesuv*" (workflow 1 result chain). In the context regular expression - knowledge resources - phonetic algorithms - language attributes - context categorisation - references - sources/material (workflow 2), the knowledge resources workflow resembles results based on a chain of "*Volcano - Vesuvius - Vesuv - Leibniz - terrae motus - letter/communication - Vesuvium - Fumarole - Solfatara*" (workflow 2 result chain).

The first connections can be found by structure, references, and regular expressions. The various Leibniz information and references in the second workflow have solely been found by phonetic algorithms. The references from English to Latin or German content has solely been possible by language attributes. In order to find further information for the result matrix even these methods would not be sufficient. Thus, the terrae motus path has been recognised using context categorisation, e.g., context keywords. With a sophisticated combination of these methods new references and new links to sources and material could be found for an improved result matrix. In this example, the term "terrae motus" has been one of the keys opening up a multitude of further information.

Material in specialised collections, for example in the European Cultural Heritage Online [22] would not be accessible due to the type and context of the material.

In the above workflow, within the chain from the stage "Leibniz" on, the content of archaeology and geosciences will not be accessible, for example the communication regarding volcanoes, earthquakes, and caves in manuscripts and letters or content of pictorial objects are not available via search engines. In this example, there is a rich contribution for the result matrix on volcanism, volcanology, and geology by various historical objects, references, and sources, especially for volcanism, Vesuvius [23], as well as earthquake related context [24], even from concept glossaries [25], manuscript collections and catalogues [26], [27] as, e.g., [28], [29], or Leibniz related copperplates [30]. For example, the "praehistoric unicorn" reconstruction [31], as well as material on geological context has not been referenced before from the objects of the knowledge resources and is not freely and publicly available as a direct reference, media or verification [32].

Therefore, with conventional search concepts, the content and any information from it will be missed within the workflow and any information will not contribute to the result matrix. Reasons for these misses can, e.g., be historical language, type of material, licensing, property and access rights. All of these being at least as important as the technical issues. Using the available features, e.g., the context categorisation from the knowledge resources it is possible to catch this information and to drastically increase the spectrum of gathering information and complementing the result matrix. The workflows and algorithms presented here can be used in order to overcome missing links in between different information pools.

Listing 11 shows an excerpt from the keyword context data of an 'Leibniz'-object.

```
...
keyword-Context: KYW :: Leibniz, Korrespondent,
 Tschirnhaus
keyword-Context: TXT :: Venedig, Neapolis, Puzzolo,
 Grotta del Cane
keyword-Context: TXT :: Neapolis, welches nach Rom und
 Venedig eine der schönsten städten Italiae ist
keyword-Context: TXT :: schwöfel bäder, schweffel
keyword-Context: KYW :: Schwefel, Solfatara, Fumarole
keyword-Context: TXT :: Neapolis, den brennenden Berg
 Vesuvium
keyword-Context: TXT :: Grotta del Cane
keyword-Context: TXT :: Neapolis, den brennenden Berg
 Vesuvium
keyword-Context: KYW DE :: Vulkanismus, Vulkanologie,
 Vesuv, Vesuvius, Vesuvium, Erdbeben, Beben
keyword-Context: KYW EN :: volcanism, volcanology,
 Vesuvius, Vesuvium, earthquake, quake
...
link-Context: LNK :: http://www.gwlb.de/Leibniz/
 Leibnizarchiv/Veroeffentlichungen/III7B.pdf
keyword-Context: TXT :: terrae motu, Sicilien
keyword-Context: KYW :: Erdbewegungen, Erdbeben,
 Vulkane, terrae motu, terra motus, Sicilien, Sizilien
...
link-Context: LNK :: http://echo.mpiwg-berlin.mpg.de
keyword-Context: KYW DE :: Nicolaus Seelaender,
 Nicolaus Seeländer, Kupferplatten, Leibniz, Leibniz
 Einhorn, Einhornhöhle b. Scharzfeld im Harz
...
link-Context: LNK :: http://www.leibnizcentral.de/
 CiXbase/gwlbhss/
keyword-Context: TXT :: 1631/1632 16xx, terra motus,
 fogelius
keyword-Context: KYW DE :: Erdbeben, Seismologie,
 Seismik, Fogel, Fogelius, Vulkan, Vesuvius, CiXbase,
 cixbase
keyword-Context: KYW EN :: earthquake, seismology,
 seismics, Fogel, Fogelius, volcano, Vesuvius, CiXbase,
  cixbase
...
link-Context: LNK :: http://www.leibnizcentral.com
keyword-Context: KYW DE :: Vulkan, Erdbeben,
 Seismologie
keyword-Context: KYW EN :: volcano, earthquake,
 seismology
```

Listing 11. Keyword context data from a 'Leibniz'-object (excerpt).

Listing 12 shows an excerpt from the keyword context data of two cave objects, which are referenced from the above object.

```
link-Context: LNK :: http://echo.mpiwg-berlin.mpg.de/
 content/copperplates
keywords-Context: KYW :: Leibniz, Nicolaus Seeländer,
 Kupferstichplatte, Copperplate, Baumannshöhle
link-Context: LNK :: http://echo.mpiwg-berlin.mpg.de/
 content/copperplates
keywords-Context: KYW :: Leibniz, Nicolaus Seeländer,
 Kupferstichplatte, Copperplate, Einhornhöhle, Harz
```

Listing 12. Keyword context data from cave objects (excerpt).

For finding these, the context descriptions have been evaluated [30], [22]. An example for the context description for one of these is shown in Listing 13.

```
Kupferstichplatten
Titel:  K 220 Einhorn und versteinerter Zahn
Beschriftung:   Tab. XII; Dens animalis marini Tidae
 prope Stederburgum e colle limoso effossi. Figura
 Sceleti prope Qvedlinburgum effossi.
Stecher:        Seeländer [signiert: N. Seelaender sc.]
Format:         318x196 mm
Bemerkung:      Abzug unter cua stark beschädigt. -
 Liste 1727, Nr. 23; Liste 1729a, Nr. 10. Abzug (ohne
 Tafelnummer) auch in Noviss. 56: IV,3, Bl. 12. Lt.
 Manuskript XXIII, 23b, Bl. 57' u. 57a, sollte dies
 ursprünglich Tafel X sein.
Abdruck:        Leibniz, Protogaea, Taf. XII, Text dazu
  S. 64 [über den Fund bei Quedlinburg]:Testis rei est
 Otto Gerikius, Magdeburgensis Consul, qui nostram
 aetatatem novis inventis illustravit [...] Gerikius
 igitur libro de vacuo edito, per occasionem narrat,
 repertum Sceleton unicornis in posteriore corporis
 parte, ut bruta solent, reclinatum, capite vero sursum
  levato, ante frontem gerens longe extensum cornu
 quinque fere ulnarum, crassitie cruris humani, sed
 proportione quadam decrescens. Ignorantia fossorum
 contritum particulatimque extractum est, postremo
 cornu cum capite et aliquibus costis, et spina dorsi
 atque ossibus Principi Abbatissae loci allata fuere.
 Eadem ad me perscripta sunt; additaque est figura,
 quam subiicere non alienum erit. [Zusatz im Manuskript
 , nicht im Druck:] Simile ingens animal Tidae prope
 Stederburgum nuper repertum est in monte a limo de
 cujus quodam immania ossa apud me sunt.  ;
Nachgestaltung:         Nachstich in Leibniz, Opera
 omnia, studio L. Dutens, 1768. - Wallmann, Abhandlung
 von den schätzbaren Alterthümern zu Quedlinburg, 1776,
  Tafel S. 39.  ;
Literatur:      Achim Rost, Das fabelhafte Einhorn. In:
  Die Welt im leeren Raum, 2002, S. 120-132. Vgl. dort
 auch S. 376 u. 378.
Signatur:       cup 4048
Signatur(Abzug):        cua 3203
```

Listing 13. Example for evaluated context description.

IX. Computation and Parallelisation

The computation time for about 100000 objects is about 20 seconds on one processor. As per request it is necessary to have several runs, for several references, this add up to about 10 minutes even for a simple object if done linear. Most of these processes can be done in parallel but due to the complexity of the knowledge content and the flexibility implemented thereof implemented for the knowledge resources it is not possible to have a general algorithm and type of parallelisation. The basic types of workflows used with object extraction are:

1) Linear workflows do not benefit from parallelisation inside the workflow. However, if a large number of comparable operations have to be executed, the overall application will benefit from a more or less loosely coupled parallelisation of these operations. The efficiency depends on the application using the results and triggering the events for the operations.
2) Parallel workflows can benefit from a parallelisation inside the workflow. This can, for example, result from operations inside the workflow that have to use persistent as well as volatile information processing. A simple case is a workflow based on a regular pattern expression on classified object groups using

homophones for finding additional object identities. In this case, the phonetic calculations can be done "on the fly", for finding the homophones in parallel for all objects as soon as they are delivered by the regular expression pattern search.
3) Partially parallel workflows will combine both linear and parallel sequences in their workflow.

Therefore, the degree of parallelisation depends on the height of the level of the implementation. The integration of knowledge resource structure, classification, and algorithms does provide large benefits on the result matrix:

- Long-term sustainable knowledge base,
- Improved Quality of Results,
- Improved Quality of Data,
- Maximum flexibility.

X. Integrated information and computing

The integration issues of information, communication, and computing are well understood [2], [33], [34] from the "collaboration house" framework [1] integrating information and scientific computing.

A. *Collaboration and multi-disciplinary workflow*

Based on the collaboration framework the IICS enables to collaborate on disciplines, services, and resources and operational level. It allows disciplines to participate on multi-disciplinary topics for building Information Systems and to use scientific supercomputing resources for computing, processing, and storage, even with interactive and dynamical components [35]. The screenshot (Figure 1) illustrates some features, as with Active Source, computed and filtered views, LX information, and aerial site photographs, e.g., from Google Maps. Many general aspects of dynamical use of information systems and scientific computing have been analysed with the collaboration house case studies.

B. *Integrative and synergetic effects*

With IICS we do have integrative as well as synergetic effects from the participating disciplines. For example, the Roman city of Altinum, next to Venice, Italy, would not have been re-discovered without the combination of archaeological information, aerial photographs, satellite images, and digital terrain models [36]. Even in unorganised circumstances, like with this discovery, the multi-disciplinary cooperation can lead to success. The more we need an integrated information system approach for "disciplines on demand" in order to improve the collaboration and the sustainability of results.

On the other hand we have synergetic effects with the same scenario of archaeology and geosciences, too, the research does have benefits for archaeology and geosciences as the collection of information from archaeological probing will help to describe the underground, which is of immense importance for the future of the area [37] and its attractiveness [38].

XI. Archaeological Information Systems

Anyway, there should be a principle solution, considering the hardware and software if so individually available, without restructuring complex data all the time when migrating to different architectures or to be prepared for future resources.

A. *Archaeology and geosciences*

So, in case of Archaeological Information Systems (AIS), for advanced Archaeological IICS, cultural heritage, and geoscientific information, and computing systems, there is a strong need for integration and documentation of different data and information with advanced scientific computing, e.g., but not limited to:

- Object, site, artifact, spatial, multi-medial, photographical, textual, properties, sources, referencial information.
- Landscape and environmental information, spatial, photographical information.
- Geophysical information, geological information.
- Event information.

Important aspects with all this information are the distribution analysis and spatial mapping. With dynamical information systems for this scenario the components must enable to weave n-dimensional topics in time, use archaeological information in education, implement n-dimensional documentation, integrate sketch mapping, provide support by multi-disciplinary referencing and documentation, discovery planning, structural analysis, multi-medial referencing.

B. *Creating metadata for documentation and computing*

It will need a number of metadata types, depending from the variable type of content, describing all kind of relevant information regarding the data and the use of this data [39]. Some important groups are category, source, batch-System, OS version and implementation, libraries, information on conversion, virtualisation environment, and automation.

Currently only a few projects in some disciplines have worked on long-term content issues [40], [41], [42], [43], [44]. Commonly only three categories are relevant to archaeological projects, project level metadata (e.g., keywords, site, dates, project information, geodata), descriptive and resource level metadata (e.g., comprehensive description, documents, databases, geo-data), and file level metadata (software, hardware, accompanying files). As we saw above, from information science point of view this is by far not sufficient as there are, e.g., licensing and archiving restrictions, precision restrictions, network limitations, context of environment, hardware, and software, hardware restrictions, tools and library limitations and implementation specifics.

The long-term aspects for big heterogeneous data hold very difficult and complex challenges as big data storage facilities [45], for users there are, e.g., free public access and long-term operational issues, for context provisioning huge amount of work have to be done, e.g., handling licensing, archiving, context, hardware availability and many more.

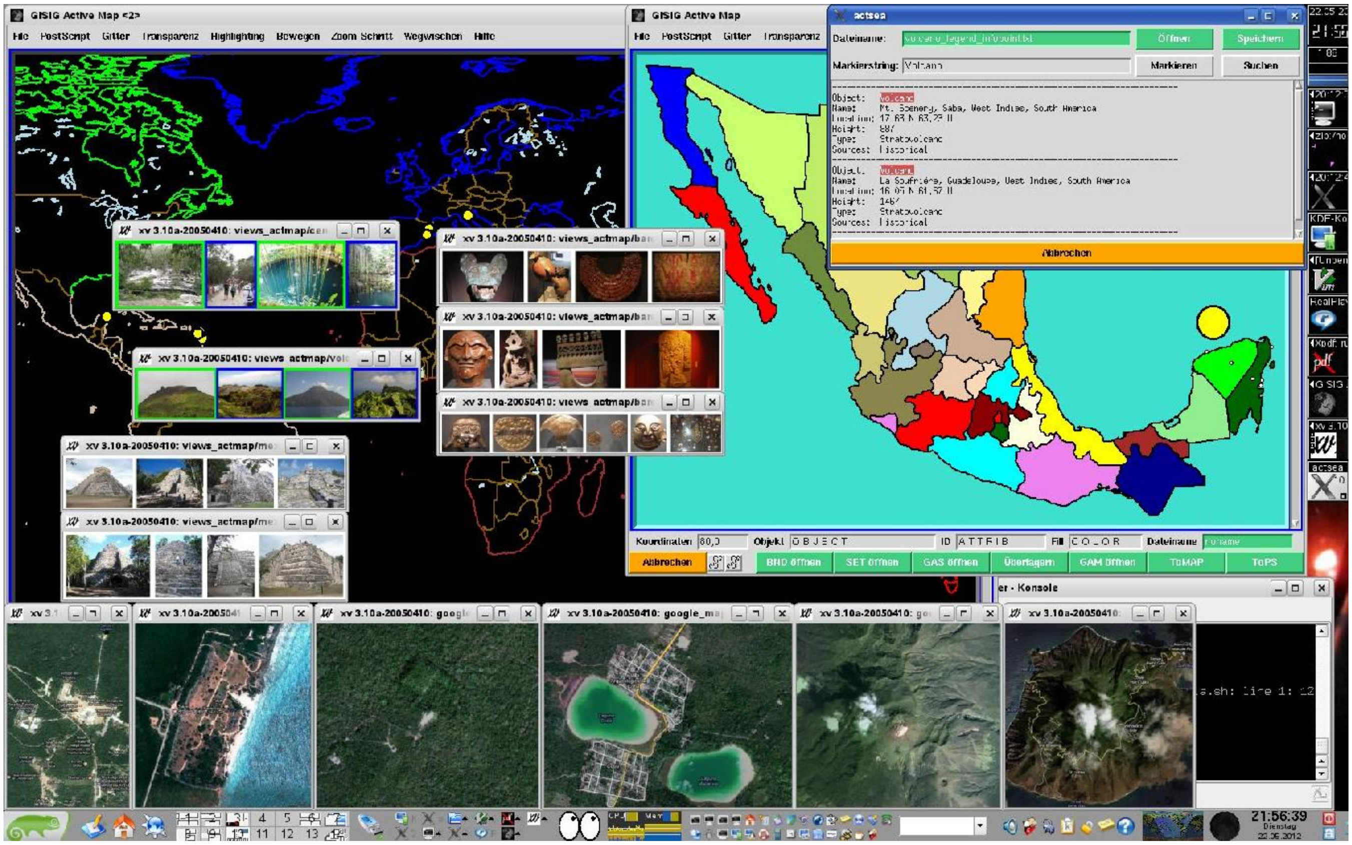

Figure 1. Dynamical use of information systems and scientific computing with multi-disciplinary and universal knowledge resources [1].

XII. Implementation of components

A. Targets and means

The main target categories and means of information to be addressed are interdisciplinary, multi-disciplinary, intercultural, functional, application, and context information. The main functional targets with IICS are integrative knowledge, education, technological glue, linking isolated samples and knowledge databases, language and transcription databases, classified Points on Interest (POI), InfoPoints, multimedial information. The organisational means are commonly grouped in disciplines, services, resources and operation.

B. Information sources

All media objects used here with components and views are provided via the Archaeology Planet and Geoscience Planet components [13]. The related information, all data, and algorithm objects presented are copyright the LX Foundation Scientific Resources [13]. It provides multi-disciplinary information and data with its knowledge resources, e.g., for archaeology, geophysics, geology, environmental sciences, geoscientific processing, geoprocessing, Information Systems, philology, informatics, computing, geoinformatics, cartography.

C. Information, structure and classification

The following examples illustrate the retrieved object information, media, and sources with examples for their multi-disciplinary relations. The information is retrieved from the LX Foundation Scientific Resources [13], [2], [46] and categorised with means like UDC. Listing 14 shows an excerpt of a LX object entry used with IICS.

```
Cenote Sagrado [Geology, Spelaeology, Archaeology]:
               Cenote, Yucatán, México.
               Holy cenote in the area of Chichén Itzá.
               ...
               %%UDC:[55+56+911.2]:[902+903+904]:
                [25+930.85]"63"(7+23+24)=84/=88
               %%Location: 20.687652,-88.567674
               Syn.: Cenote Sagrada
               s. also Cenote, Chichén Itzá
```

Listing 14. Structure of object entry (LX Resources, excerpt).

Listing 15 shows a classification set of UDC samples used with the knowledge resources and IICS.

```
UDC:[902+903+904]:[25+930.85]"63"(7)(093)=84/=88
UDC:[902+903+904]:[930.85]"63"(23)(7):(4)=84/=88
UDC:[55+56+911.2]:[902+903+904]:[25+930.85]"63"
  (7+23+24)=84/=88
UDC:[25+930.85]:[902]"63"(7)(093)=84/=88
UDC:[911.2+55+56]:[57+930.85]:[902+903+904]"63"
  (7+23+24)=84/=88
UDC:[911.2+55]:[57+930.85]:[902]"63"(7+23+24)=84/=88
```

Listing 15. Classification set (UDC samples, excerpt).

The classification deployed for documentation [47] must be able to describe any object with any relation, structure, and level of detail. Objects include any media, textual documents, illustrations, photos, maps, videos, sound recordings, as well as realia, physical objects such as museum objects. A suitable background classification is, e.g., the UDC. The objects use preliminary classifications for multi-disciplinary content. Standardised operations with UDC are, e.g., addition ("+"), consecutive extension ("/"), relation (":"), subgrouping ("[]"), non-UDC notation ("*"), alphabetic extension ("A-Z"), besides place, time, nationality, language, form, and characteristics.

D. Communication and computing

The central component groups for bringing multi-disciplinary information systems into practice are IICS and documentation of objects, structure, and references. Listing 16 shows an example of a dynamical dataset from an Active Source [35] component provisioning information services.

```
#BCMT--------------------------------------------------
###EN \gisigsnip{Object Data: Country Mexico}
#ECMT--------------------------------------------------
proc create_country_mexico {} {
global w
$w create polygon 0.938583i 0.354331i 2.055118i ...
...
proc create_country_mexico_autoevents {} {
global w
$w bind legend_infopoint <Any-Enter> {set killatleave [
 exec ./mexico_legend_infopoint_viewall.sh $op_parallel
  ] }
$w bind legend_infopoint <Any-Leave> {exec ./
 mexico_legend_infopoint_kaxv.sh }
$w bind tulum <Any-Enter> {set killatleave [exec
 $appl_image_viewer -geometry +800+400 ./
 mexico_site_name_tulum_temple.jpg $op_parallel ] }
$w bind tulum <Any-Leave> {exec kill -9 $killatleave }
} ...
```

Listing 16. Dynamical data set of Active Source component.

Batch and interactive features are integrated with Active Source event management [35], e.g., allowing structure and UDC based filtering. Computing interfaces can carry any interactive or batch job description. Taking a look onto different batch and scheduling environments one can see large differences in capabilities, handling environments and architectures. In the last years, experiences have been gained in simple features for different environments for High Throughput Computing like Condor, workload schedulers like LoadLeveler and Grid Engine, and batch environments like Moab / Torque.

XIII. Resulting implementation in practice

A. Scientific documentation

Scientific documentation is an essential part of a Universal IICS (UIICS), revealing associations and relations and gaining new insight. Handling the available information does provide transparent how puzzle pieces of a scientific context do fit, e.g., not only that terms like Bronze Age, Ice Age, Stone Age are only regional but in quantity and quality how the transitions and distributions in space and time are. Information on objects, archiving, analysis, documentation, sources and so on will be provided as available with the dimension space. Besides the dynamical features the objects carry information, e.g., references, links, tags, and activities.

B. Dimension space

The information matrix spans a multi-dimensional space (Table I). It illustrates the multi-faceted topic dimension containing important cognitive information for disciplines and applications. Examples of multi-disciplinary information in archaeological context are stony and mineral composition, e.g., of dead freight or ballast in ship wrecks, mineral material in teeth, fingerprints of metals used in artifacts, and genetic material of biological remains. Further there exists a "vertical" multi-dimensional space to this information matrix, carrying complementary information, e.g., color, pattern, material, form, sound, letters, characters, writing, and so on. The documentation can handle the holistic multi-dimensional space, so we can flatten the views with available interfaces to three or four dimensional representations.

Table I
Dimensions of the information matrix (excerpt).

Dimension	*Meaning, Examples*
Time	Chronology
Topic	Disciplines
	Purpose (tools, pottery, weapons, technology, architecture, inscriptions, sculpture, jewellery)
	Culture (civilisation, ethnology, groups, etymology)
	Infrastructure (streets, pathways, routes)
	Environment (land, sea, geology, volcanology, speleology, hydrogeology, astronomy, physics, climatology)
	Genealogy (historical, mythological documentation)
	Genetics (relationship, migration, human, plants)
	Biology (plants, agriculture, microorganisms)
	Trade (mobility, cultural contacts, travel)
Depth	Underground, subterranean
Site	Areal distribution, region
...	...
Data	Resources level, virtualisation

The dimensions are not layers in any way so it would contradict to percept their documentation with integrated systems in data or software layers. With these IICS we are facing a multi-dimensional volume, like multi-dimensional "potato shapes" of knowledge objects. Layer concepts are often used with cartographic or mapping applications but these products are infeasible for handling complex cognitive context.

C. IICS dimension view

As with the structure the communication and compute processes are getting resource intensive, the available storage and compute resources are used with the IICS. The following small example shows an excerpt of a tabulated dimension view (Table II). The last column shows if an object is deposited on site (O) or distributed (D) and if additional media is available and referenced. The table shows if a storage or and additional compute request has been necessary for the resulting object or media. Information is given if primarily a storage request (S) for persistant media or a compute request (C) deploying High End Computing resources is dynamically used for creating the appropriate information.

Table II
DIMENSION VIEW WITH ARCHAEOLOGICAL IICS (EXCERPT).

Topic	*Purpose / Environment / Infrastructure*	*Ref.*
Egypt	Architecture	
Rome	Architecture	
Catalonia	Architecture	
	Monument de Colom, Port, Barcelona, Spain	OC
Maya	Architecture	
	Kukulkán Pyramid, Chichén Itzá, Yucatán, México	OC
	Nohoch Mul Pyramid, Cobá, Yucatán, México	OC
	El Meco Pyramid, Yucatán, México	OC
	El Rey Pyramid, Cancún, Yucatán, México	OC
	Pelote area, Cobá, Yucatán, México	OS
	Pok ta Pok, Cancún, Yucatán, México	OS
	Templo del Alacran, Cancún, Yucatán, México	OS
	Port, Tulúm, Yucatán, México	OC
	Infrastructure	
	Sacbé, Chichén Itzá, Yucatán, México	OS
	Sculpture	
	Diving God & T. Pinturas, Tulúm, Yucatán, México	OC
	Diving God, Cobá, Yucatán, México	OC
Precolombian	Architecture	
Caribbean	Environment (volcanology, geology, hydrogeology)	
	La Soufrière volcano, Guadeloupe, F.W.I.	OC
	Mt. Scenery volcano, Saba, D.W.I.	OC
	Cenote Sagrado, Chichén Itzá, Yucatán, México	OC
	Ik Kil Cenote, Yucatán, México	OC
Arawak	Architecture	
Prehistory	Architecture	
Topic:	architecture mythology environment infrastructure	
Entity:	Object Location: O On site, D Distributed; Object Media: C Compute, S Storage.	
Compute:	CONNECT REFERTO-TOPIC REFERTO-SPATIAL VIEW-TO VIEW-FROM	

The following examples explain views from disciplines and topics (Figure 1) as computed and filtered with the IICS, using photo media samples (media samples © C.-P. Rückemann, 2012, 2013). It must be emphasised that the applications can provide any type of objects, high resolution media, and detailed information. The first view (Figure 2) is a simple example from the above table for an excerpt of the computed class of regional pyramid object representations (Yucatán Peninsula, provinces Yucatán and Quintana Roo).

Figure 2. Object SAMPLE – regional pyramid of Maya, Yucatán, México.

Figure 3 illustrates the computed objects for the above REFERTO-TOPIC and REFERTO-SPACE chain classification, e.g., here via UDC "`(7):(4)`" relation.

Figure 3. Cross-purpose REFERTO – Diving god, Tulúm, Colom.

Besides that, viewing directions can be referred, e.g., "view to", "view from", "detail" as shown with a VIEW example (Figure 4) for the above selection with UDC "`(23)`", "`(24)`".

Figure 4. In-purpose: VIEW-TO VIEW-FROM – Volcanoes and Cenotes.

D. Topic view and object representation

The following sample excerpt tabulates a topic view (Table III) and shows the computed object representation (Figure 5) for an in-topic CONNECT example. From the eight samples of Chichén Itzá shown, the Sacbé pathway connects the Kukulkán Pyramid with the Cenote Sagrado. The table shows a sample of referred (Geo) information.

Table III
TOPIC VIEW WITH ARCHAEOLOGICAL IICS (EXAMPLE, CHICHÉN ITZÁ).

Site	*Topic / Purpose*	*Selected: Geo*	*Ref*
Chichén Itzá			
	Kukulkán Pyramid, El Castillo	Limestone	OC
	Sacbé	Limestone	OC
	Cenote Sagrado	Doline, hydrology	OC
	Jaguar temple		OS
	Tzompantli		OS
	Temple of the warriors		OS
	Caracol		OS
	Chac temple		OS

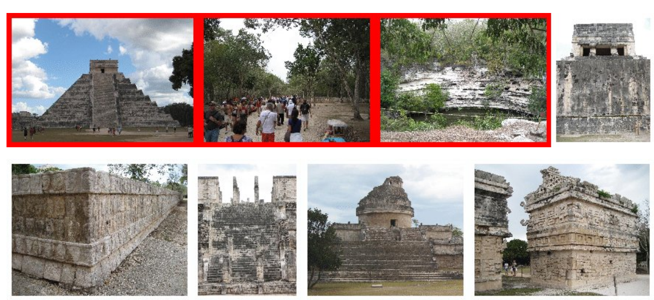

Figure 5. In-topic CONNECT – Kukulkán, Cenote, connected by Sacbé.

As Figure 1 showed, the objects resulting from the computation can contain any additional attributes, e.g., geo-referenced relations for further application within spatial context or multi-disciplinary analysis and evaluation.

E. Object space grouping

The objects are linked by relations in the n-dimensional object space. The slices with a selected number of dimensions carry the common information, e.g., "Stone Age flint arrow heads" in a specific area. It is essential not to sort objects into layers within a database-like structure. So vectors and relations can help to represent their nature in a more natural way. The views, even traditional layered ones, are created from these by appropriate components. The following figures illustrate structure and references for collections, context, and integration of multi-disciplinary information: museum topical collection (Figure 6), context of amphores (Figure 7), and geology information (Figure 8).

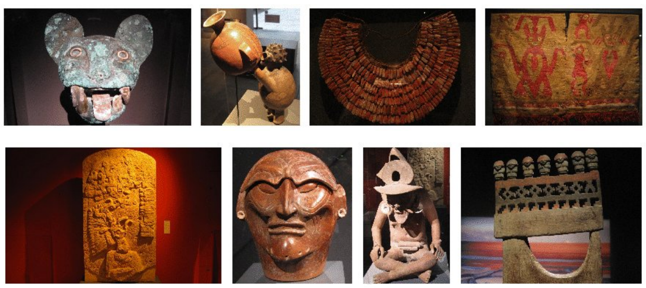

Figure 6. Sample COLLECTION – Precolombian Museum.

Figure 7. Sample CONTEXT – Pottery (amphores).

Figure 8. Sample DISCIPLINE – Geology (Caribbean limestone and tuff).

XIV. Digital archaeological library examples

In combination with the above shown features, objects in digital archaeological libraries have been enriched with various information, e.g., on museum, library information, archives, network information, mapping services, locations and Points Of Interest (POI).

Due to the knowledge resources organisation, the objects can be used in references as well as in the cache for interactive components at any stage within the workflow process. Combining the structure and classification with the silken selection algorithms leads to very flexible, multi-disciplinary interfaces.

Each group of digital images shows the result matrix from a selection process. The following figures illustrate resulting objects from the digital library of the LX knowledge resources with multi-disciplinary background, in these examples regarding material, function, and model or reconstruction purposes.

Selecting archaeological objects from Central and Southern America, ancient art, and consisting of Gold (`UDC:902+(7),(8)+700.32+546.59`) results in a subset from the gold objects collection (Figure 9).

Figure 9. Sample COLLECTION – Jewelery + material: gold.

Selecting archaeological objects, ancient art, and being part of the collier collection (`UDC:902+700.32`) results in a subset from the jewelery collection (Figure 10).

Figure 10. Sample COLLECTION – Jewelery + function: collier.

The result shows, that objects can be member of any number of collections and result matrices, as compared to the result from the museum topical collection (Figure 6). Selecting archaeological objects, watercraft engineering, marine engineering, boats, ships, boat building, ship building, and being models having origin from ancient Egypt and the Mediterranean (`UDC:902+629.5+(32),(37),(38)`) results in a subset from the ship model collection (Figure 11).

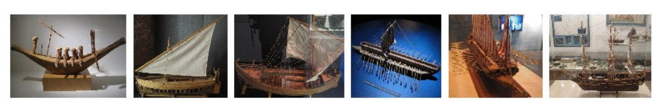

Figure 11. Sample COLLECTION – Ship + type: model.

Any silken criteria can be used for the transliteration, transcriptions and other content and context. So if not limited to fixed criteria, the resulting matrix is highly dynamical and supports a flexible modelling of the relations and explicit and implicit references within the available material.

XV. Conclusion and future work

It has been shown how long-term knowledge resources have been created and used for more than twenty-five years considering content and context with sophisticated workflows implementing various technology over the years.

The knowledge resources have proven to provide a universal way of describing multi-disciplinary objects, expressing relations between any kind of objects and data, e.g., from archaeology, geosciences, and natural sciences as well as defining workflows for calculation and computation for application components. Systematically structuring, classification, as well as soft 'silken' criteria with LX and UDC support have provided efficient and economic means for using Information System components and supercomputing resources. With these, the solution scales, e.g., regarding references, resolution, and view arrangements even with big data scenarios and parallel computing resources. The concept can be transferred to numerous applications in a very flexible way and has shown to be most sustainable.

The successful integration of IICS components and advanced scientific computing based on structured information and faceted classification of objects has provided a very flexible and extensible solution for the implementation of Archaeological Information Systems.

It has been demonstrated with the case studies that Archaeological IICS can provide advanced multi-disciplinary information as from archaeology and geosciences by means of High End Computing resources.

The basic architecture has been created using the collaboration house framework, long-term documentation and classification of objects, flexible algorithms, workflows and Active Source components. As shown with the examples, any kind of computing request, e.g., discovery, data retrieval, visualisation, and processing, can be done from the application components accessing the knowledge resources. Computing interfaces can carry any interactive or batch job description. Anyhow, the hardware and system resources have to be configured appropriately for a use with the workflow. For future applications a kind of "tooth system" for long-term documentation and algorithms for use with IICS and the exploitation of supercomputing resources will be developed. Besides this, it is intended to further extend the content spectrum of the knowledge resources.

Acknowledgements

We are grateful to all national and international academic, industry, and business partners in the GEXI cooperations and the Science and High Performance Supercomputing Centre (SHPSC) for long-term support of collaborative research and the LX Project for providing suitable resources. Many thanks to the scientific colleagues at the Leibniz Universität Hannover, the Institute for Legal Informatics (IRI), and the Westfälische Wilhelms-Universität (WWU), sharing experiences on ZIV, HLRN, Grid, and Cloud resources and for participating in fruitful case studies as well as the participants of the INFOCOMP and DigitalWorld conferences as well as the postgraduate European Legal Informatics Study Programme (EULISP) for prolific discussion of scientific, legal, and technical aspects over the last years.

Thanks for excellent inspiration, support, and photo scenery go to the Saba Conservation Foundation, Saba Marine Park, and National Heritage Foundation St. Maarten (D.W.I.), National Park Guadeloupe and Museum St. Martin (F.W.I.), Instituto Nacional de Antroplogía e Historia (I.N.A.H.), Mexicó for providing access to the sites of Chichén Itzá, Cobá, Tulúm, El Meco, El Rey, and many more as well as to the Eco-Parc Xel Há, México, and especially to Ms. Maureen Felix (Consejo de Promoción Turística de México, CPTM) for her excellent support, the Museu Barbier-Mueller d'Art Precolombí and Museu Egipci Barcelona, Museu Urbana, Valencia, Spain, as well as Canon for the photo equipment.

References

[1] 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. and Dini, P. and Hommel, W. and Pankowska, M. and Schubert, L. (eds.), ISBN: 978-1-61208-226-4, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2012_3_10_10012 [accessed: 2013-06-09].

[2] C.-P. Rückemann, *Queueing Aspects of Integrated Information and Computing Systems in Geosciences and Natural Sciences*. InTech, 2011, pp. 1–26, Chapter 1, in: Advances in Data, Methods, Models and Their Applications in Geoscience, 336 pages, ISBN-13: 978-953-307-737-6, DOI: 10.5772/29337, OCLC: 793915638, DOI: http://dx.doi.org/10.5772/29337 [accessed: 2013-05-26].

[3] C.-P. Rückemann, "Implementation of Integrated Systems and Resources for Information and Computing," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2011), October 23–29, 2011, Barcelona, Spain*, 2011, pp. 1–7, ISBN: 978-1-61208-009-3, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2011_1_10_10002 [accessed: 2013-05-26].

[4] "Geo Exploration and Information (GEXI)," 1996, 1999, 2010, 2013, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/index.html#GEXI (Information) [accessed: 2013-05-26].

[5] L. Yin, S.-L. Shaw, D. Wang, E. A. Carr, M. W. Berry, L. J. Gross, and E. J. Comiskey, "A framework of integrating GIS and parallel computing for spatial control problems - a case study of wildfire control," *IJGIS,* ISSN: 1365-8816, DOI: 10.1080/13658816.2011.609487, pp. 1–21, 2011.

[6] National Park Service, "National Register of Historic Places Official Website, Part of the National Park Service (NPS)," 2013, NPS, URL: http://www.nps.gov/nr [accessed: 2013-05-26].

[7] "North American Database of Archaeological Geophysics (NADAG)," 2013, University of Arkansas, URL: http://www.cast.uark.edu/nadag/ [accessed: 2013-05-26].

[8] "Center for Advanced Spatial Technologies (CAST)," 2013, University of Arkansas, URL: http://www.cast.uark.edu/ [accessed: 2013-05-26].

[9] "Archaeology Data Service (ADS)," 2013, URL: http://archaeologydataservice.ac.uk/ [accessed: 2013-05-26].

[10] "Center for Digital Antiquity," 2013, Arizona State Univ., URL: http://www.digitalantiquity.org/ [accessed: 2013-05-26].

[11] "The Digital Archaeological Record (tDAR)," 2013, URL: http://www.tdar.org [accessed: 2013-05-26].

[12] IBM, "City Government and IBM Close Partnership to Make Rio de Janeiro a Smarter City," *IBM News room - 2010-12-27, USA*, 2012, URL: http://www-03.ibm.com/press/us/en/pressrelease/33303.wss [accessed: 2012-03-18].

[13] "LX-Project," 2013, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX (Information) [accessed: 2013-05-26].

[14] "Universal Decimal Classification Consortium (UDCC)," 2013, URL: http://www.udcc.org [accessed: 2013-02-10].

[15] R. C. Russel and M. K. Odell, "U.S. patent 1261167," 1918, (Soundex algorithm), patent issued 1918-04-02.

[16] D. E. Knuth, *The Art of Computer Programming: Sorting and Searching*. Addison-Wesley, 1973, vol. 3, ISBN: 978-0-201-03803-3, OCLC: 39472999.

[17] National Archives and Records Administration, "The Soundex Indexing System," 2007, 2007-05-30, URL: http://www.archives.gov/research/census/soundex.html [accessed: 2013-05-26].

[18] M. Stok, "Perl, Soundex.pm, Soundex Perl Port," 1994, (code after Donald E. Knuth).

[19] "LX SNDX, a Soundex Module Concept for Knowledge Resources," *LX-Project Consortium Technical Report*, 2013, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX (Information) [accessed: 2013-05-26].

[20] E. Rempel, "tcllib, soundex.tcl, Soundex Tcl Port," 1998, (code after Donald E. Knuth).

[21] A. Kupries, "tcllib, soundex.tcl, Soundex Tcl Port Documentation," 2003, (code after Donald E. Knuth).

[22] Max Planck Institute for the History of Science, Max-Planck Institut für Wissenschaftsgeschichte, "European Cultural Heritage Online (ECHO)," 2013, Berlin, URL: http://echo.mpiwg-berlin.mpg.de/ [accessed: 2013-05-26].

[23] E. W. von Tschirnhaus, "Brief (Letter), Ehrenfried Walther von Tschirnhaus an Leibniz 17.IV.1677," pp. 59–73, 1987, Gottfried Wilhelm Leibniz, Sämtliche Schriften und Briefe, Mathematischer, naturwissenschaftlicher und technischer Briefwechsel dritte Reihe, zweiter Band, 1667 – 1679, Leibniz-Archiv der Niedersächsischen Landesbibliothek Hannover, Akademie-Verlag Berlin, 1987, herausgegeben unter Aufsicht der Akademie der Wissenschaften in Göttingen; Akademie der Wissenschaften der DDR.

[24] G. F. von Franckenau, "Brief (Letter), Georg Franck von Franckenau an Leibniz 18. (28.) September 1697, Schloss Frederiksborg, 18. (28.) September 1697," pp. 568–569, Gottfried Wilhelm Leibniz Bibliothek (GWLB), Leibniz-Archiv der Niedersächsischen Landesbibliothek Hannover, URL: http://www.gwlb.de/Leibniz/Leibnizarchiv/Veroeffentlichungen/III7B.pdf [accessed: 2013-05-26] .

[25] Berlin-Brandenburgische Akademie der Wissenschaften, "Leibniz Reihe VIII," 2013, Glossary, Concepts, BBAW, Berlin, URL: http://leibnizviii.bbaw.de/glossary/concepts/ [accessed: 2013-05-26] (concepts glossary), URL: http://leibnizviii.bbaw.de/Leibniz_Reihe_8/Aus+Otto+von+Guericke,+Experimenta+nova/LH035,14,02_091v/index.html [accessed: 2013-05-26] (transcription), URL: http://leibnizviii.bbaw.de/pdf/Aus+Otto+von+Guericke,+Experimenta+nova/LH035.14,02_091v/LH035,14!02_091+va.png [accessed: 2013-05-26] (scan).

[26] Gottfried Wilhelm Leibniz Bibliothek (GWLB), Niedersächsische Landesbibliothek, "GWLB Handschriften," 2013, hannover, URL: http://www.leibnizcentral.de/CiXbase/gwlbhss/ [accessed: 2013-05-26].

[27] "LeibnizCentral," 2013, URL: http://www.leibnizcentral.com/ [accessed: 2013-02-10].

[28] M. Fogel, "Brieffragmente (Letter fragments) about 16xx, Historici Pragmatici universal, Terrae motus, Physica," manuscript ID: 00016293, Source: Gottfried Wilhelm Leibniz Bibliothek (GWLB), Niedersächsische Landesbibliothek, GWLB Handschriften, Hannover, URL: http://www.leibnizcentral.de/CiXbase/gwlbhss/ [accessed: 2013-05-26].

[29] M. Fogel, "Brieffragmente (Letter fragments) about 16xx, Terrae Motus in Nova Francia," manuscript ID: 00016278, Source: Gottfried Wilhelm Leibniz Bibliothek (GWLB), Niedersächsische Landesbibliothek, GWLB Handschriften, Hannover, URL: http://www.leibnizcentral.de/CiXbase/gwlbhss/ [accessed: 2013-05-26].

[30] Gottfried Wilhelm Leibniz Bibliothek Hannover, "Collection of Copperplates," 2013, URL: http://echo.mpiwg-berlin.mpg.de/content/copperplates [accessed: 2013-05-26].

[31] N. Seeländer, "Dens animalis marini Tidae prope Stederburgum e colle limoso effossi, Figura Sceleti prope Qvedlinburgum effossi," about 1716, Copperplate, (Kupferstichplatten), printed in "Leibniz, Protogaea, Tab. XII", re-printed in "Leibniz, Opera omnia, studio L. Dutens, 1768 – Wallmann, Abhandlung von den schätzbaren Alterthümern zu Quedlinburg, 1776, Tafel S. 39", URL: http://echo.mpiwg-berlin.mpg.de/ECHOdocuView?url=/mpiwg/online/permanent/echo/copperplates/Leibniz_cup4/pageimg&start=41&pn=73&mode=imagepath [accessed: 2013-05-26].

[32] C.-P. Rückemann and B. F. S. Gersbeck-Schierholz, "Object Security and Verification for Integrated Information and Computing Systems," in *Proceedings of the Fifth International Conference on Digital Society (ICDS 2011), Proceedings of the International Conference on Technical and Legal Aspects of the e-Society (CYBERLAWS 2011), February 23–28, 2011, Gosier, Guadeloupe, France / DigitalWorld 2011*. XPS, 2011, pp. 1–6, ISBN: 978-1-61208-003-1, URL: http://www.thinkmind.org/download.php?articleid=cyberlaws_2011_1_10_70008 [accessed: 2013-05-26].

[33] C.-P. Rückemann, "Dynamical Parallel Applications on Distributed and HPC Systems," *International Journal on Advances in Software*, vol. 2, no. 2&3, pp. 172–187, 2009, ISSN: 1942-2628, LCCN: 2008212462 (Library of Congress), URL: http://www.thinkmind.org/index.php?view=article&articleid=soft_v2_n23_2009_1/ [accessed: 2013-05-26] (ThinkMind(TM) Digital Library), URL: http://www.iariajournals.org/software/soft_v2_n23_2009_paged.pdf [accessed: 2013-05-26].

[34] C.-P. Rückemann, "Legal Issues Regarding Distributed and High Performance Computing in Geosciences and Exploration," in *Proceedings of the International Conference*

on Digital Society (ICDS 2010 / CYBERLAWS 2010), February 10–16, 2010, St. Maarten, Netherlands Antilles, D.W.I. IEEE Computer Society Press, IEEE Xplore Digital Library, 2010, pp. 339–344, ISBN: 978-0-7695-3953-9, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5432414 [accessed: 2013-05-26].

[35] C.-P. Rückemann, "Beitrag zur Realisierung portabler Komponenten für Geoinformationssysteme. Ein Konzept zur ereignisgesteuerten und dynamischen Visualisierung und Aufbereitung geowissenschaftlicher Daten," Dissertation, WWU, Münster, Deutschland, 2001, 161 (xxii+139) S., Deutsche Nationalbibliothek H 2002 A 103 (German National Library), urn:nbn:de:swb:14-1011014626187-99999 (persistent URN), oai:d-nb.de/dnb/964146754 (German National Library), OCLC: 50979238, 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: 2013-05-26].

[36] J. A. Lobell, "Roman Venice Discovered," *Archaeological Institute of America, November/December 2009*, vol. 62, no. 6, 1996, URL: http://www.archaeology.org/0911/trenches/roman_venice.html [accessed: 2013-05-26].

[37] A. J. Ammerman, "Probing the Depths of Venice," *Archaeological Institute of America, July/August 1996*, vol. 49, no. 4, 1996, URL: http://www.archaeology.org/9607/abstracts/venice.html [accessed: 2013-05-26].

[38] "Venice Mobility Project - Pedestrian Modeling," *Santa Fe Complex*, 2012, February, 2012, URL: http://sfcomplex.org/2012/02/venice-mobility-project-pedestrian-modeling [accessed: 2013-05-26].

[39] C.-P. Rückemann, "Advanced Scientific Computing and Multi-Disciplinary Documentation for Geosciences and Archaeology Information," in *Proceedings of The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2013), February 24 – March 1, 2013, Nice, Cote d'Azur, French Riviera, France*. XPS Press, 2013, pp. 81–88, Rückemann, C.-P. (ed.), ISSN: 2308-393X, ISBN: 978-1-61208-251-6, URL: http://www.thinkmind.org/download.php?articleid=geoprocessing_2013_4_10_30035 [accessed: 2013-05-26], URL: http://www.iaria.org/conferences2013/ProgramGEOProcessing13.html (Program) [accessed: 2013-05-26].

[40] K. Perrin, "Archaeological Archives: Documentation, Access and Deposition. A Way Forward," *English Heritage*, 2002.

[41] D. H. Brown, "Safeguarding Archaeological Information: Procedures for minimising risk to undeposited archaeological archives," *English Heritage*, 2011, URL: http://www.english-heritage.org.uk/publications/safeguarding-archaeological-information/ [accessed: 2012-04-08].

[42] "Guides to Good Practice," 2013, ADS, URL: http://guides.archaeologydataservice.ac.uk/ [accessed: 2013-05-26].

[43] H. Eiteljorg II, K. Fernie, J. Huggett, and D. Robinson, *CAD: A Guide to Good Practice*. Archaeology Data Service, 2002, ISSN: 1463-5194, URL: http://ads.ahds.ac.uk/project/goodguides/cad/ [accessed: 2013-05-26].

[44] "Archaeological Archives Forum (AAF)," 2013, URL: http://www.britarch.ac.uk/archives/ [accessed: 2013-05-26].

[45] *DigitalWorld 2012 / GEOProcessing International Expert Panel on Challenges in Handling Large Data Volume for GEO Processing, January 31, 2012, Valencia, Spain The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012), Polytechnic University of Valencia, January 30 – February 4, 2012, Valencia, Spain*, 2012, URL: http://www.iaria.org/conferences2012/filesGEOProcessing12/GEO_2012_PANEL.pdf [2013-05-26], URL: http://www.iaria.org/conferences2012/ProgramGEOProcessing12.html (Program) [accessed: 2013-05-26].

[46] C.-P. Rückemann, *Integrated Information and Computing Systems for Advanced Cognition with Natural Sciences*. Premier Reference Source, Information Science Reference, IGI Global, 701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA, Oct. 2012, pp. 1–26, chapter I, in: Rückemann, C.-P. (ed.), Integrated Information and Computing Systems for Natural, Spatial, and Social Sciences, 543 (xxiv + 519) pages, 21 chapters, ill., ISBN-13: 978-1-4666-2190-9 (hardcover), EISBN: 978-1-4666-2191-6 (e-book), ISBN: 978-1-4666-2192-3 (print & perpetual access), DOI: 10.4018/978-1-4666-2190-9, LCCN: 2012019278 (Library of Congress), OCLC: 798809710, URL: http://www.igi-global.com/chapter/integrated-information-computing-systems-advanced/70601 [accessed: 2013-05-26], DOI: http://dx.doi.org/10.4018/978-1-4666-2190-9.ch001 [accessed: 2013-05-26].

[47] C.-P. Rückemann, "Integrating Information Systems and Scientific Computing," *International Journal on Advances in Systems and Measurements*, vol. 5, no. 3&4, pp. 113–127, 2012, ISSN: 1942-261x, LCCN: 2008212470 (Library of Congress), URL: http://www.thinkmind.org/index.php?view=article&articleid=sysmea_v5_n34_2012_3/ [accessed: 2013-05-26] (ThinkMind(TM) Digital Library), URL: http://www.iariajournals.org/systems_and_measurements/sysmea_v5_n34_2012_paged.pdf [accessed: 2013-06-09].