Requirements Gathering and Modeling of Domain-Specific Digital Libraries with the 5S Framework: An Archaeological Case Study with ETANA

Rao Shen[1], Marcos André Gonçalves[2], Weiguo Fan[1], Edward Fox[1]

[1] Digital Library Research Laboratory, Virginia Tech, USA
[2] Department of Computer Science, Federal University of Minas Gerais (UFMG), Brazil
{rshen, mgoncalv, wfan, fox}@vt.edu

Abstract. Requirements gathering and conceptual modeling are essential for the customization of digital libraries (DLs), to help attend the needs of target communities. In this paper, we show how to apply the 5S (Streams, Structures, Spaces, Scenarios, and Societies) formal framework to support both tasks. The intuitive nature of the framework allows for easy and systematic requirements analysis, while its formal nature ensures the precision and correctness required for semi-automatic DL generation. More specifically, we show how 5S can help us define a domain-specific DL metamodel in the field of archaeology. Finally, an archaeological DL case study (from the ETANA project) yields informal and formal descriptions of two DL models (instances of the metamodel).

1 Introduction

The construction of any digital library (DL) involves a number of decisions covering: 1) which types of multimedia content will be supported by the DL; 2) how the stored information is organized and structured; 3) which are the target communities; and 4) which services and capabilities will be provided to them [3]. The process of formally assembling such decisions and representing them in a format useful for processing by a DL system involves both requirements gathering and analytical modeling or design.

Modern software engineering has encouraged the use of formal methods, with mathematically defined syntax and semantics, to support such tasks. Formal methods and frameworks can support specification of (most of the parts of) complex systems such as DLs, while also promoting rigor and correctness. This paper focuses on the application of the 5S (Streams, Structures, Spaces, Scenarios, and Societies) formal framework [2] in the support of these tasks. More specifically, we show how 5S can help us document complex requirements and can support the modeling of domain-specific digital libraries, as is illustrated with a case study from the field of archaeology.

The rest of this paper is organized as follows. Section 2 provides a brief background on the 5S framework. Section 3 informally discusses requirements of archaeological DLs according to 5S. Section 4 builds on the previous section to present a formal

archaeological DL metamodel. Section 5 presents a two-part case study illustrating the methodology and models. Section 6 concludes the paper and outlines future work.

2 Background on the 5S Framework

In [4] we presented a formal framework for the DL field, summarized in Figure 1. We defined "minimal digital library" (defn. 24 of [4], shown at the bottom right). Figure 1 illustrates the supporting layers of definitions: mathematical foundations (e.g., graphs, sequences, functions), the 5 Ss (Streams, Structures, Spaces, Scenarios, and Societies), and key concepts of a DL (e.g., digital object, collection). Arrows represent dependencies, indicating that a concept is formally defined in terms of previously defined concepts that point to it, "digital library (minimal)" being the highest level concept.
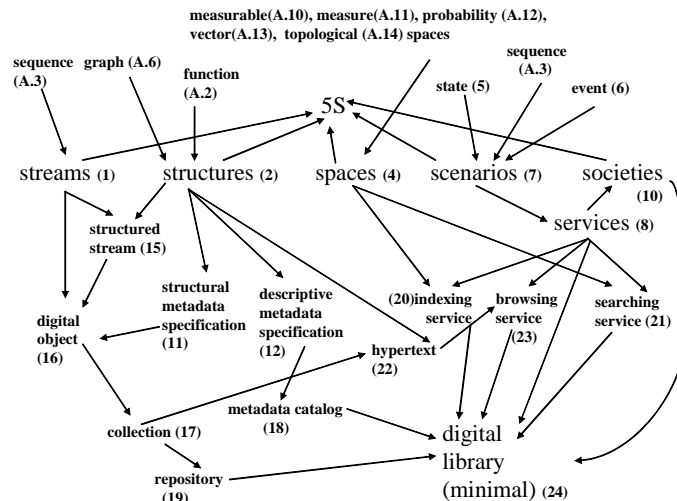

Fig. 1. 5S definitional structure (from [4]).

3 Archaeological Digital Libraries: a 5S-Based Informal View

This section shows how 5S can be used to analyze the requirements of domain-specific DLs. More specifically, it informally describes the archaeological domain, and therefore archaeological DLs (ArchDLs), in the light of the 5S framework.

1.   Societies
Societies can be groups of humans as well as hardware and software components. Examples of human societies in ArchDLs include archaeologists (in academic institutes, fieldwork settings, excavation units, or local and national government bodies), the general public (e.g., educators, learners), and those who lived in historic and pre-historic societies. There also are societies of project directors, field staff (responsible

for the actual work of excavation), technical staff (consisting of photographers, technical illustrators, and their assistants), and camp staff (including camp managers, registrars, and tool stewards). Since archaeology is a multi-disciplinary subject, drawing on a wide range of skills and specialties, from the arts and humanities through to the biological and physical sciences, societies of specialists (e.g., in geology, anthropology, lithics, ceramics, faunal and floral remains, remote sensing) are involved in ArchDLs. Societies follow certain rules and their members play particular roles. Members of societies have activities and relationships (e.g., specialists serve to assist and advise the varying field and laboratory staffs regarding field problems and other matters related to their special skills and interests). Because archaeologists in diverse countries follow different laws and customs, a number of ethical and freedom-related issues arise in connection with ArchDLs. Examples include: Who owns the finds? Where should they be preserved? What nationality and ethnicity do they represent? Who has publication rights? To address these issues, and to support the variety of needs of interested societies, DL designers have planned for numerous scenarios.

2.   Scenarios

A scenario is often defined as a description of interactions between a human user and a computerized system. Scenarios also can describe interactions among software modules (as in [4]) or among humans. Further, describing scientific processes (hypothesizing, observing, recording, testing, analyzing, and drawing conclusions – used during any archaeological study) as scenarios can help with comprehending specific ArchDL phenomena, and with requirements elicitation and specification generation.

Digital recording as an archaeological process to facilitate information gathering occurs in two stages, the planning stage and the excavation stage. Remote sensing, fieldwalking, field surveys, building surveys, consulting historical and other documentary sources, and managing the sites and monuments (and related records) maintained by local and national government bodies may be involved in the planning stage. During excavation, detailed information is recorded, including for each layer of soil, and for features such as pole holes, pits, and ditches. Data about each artifact is recorded together with information about its exact find spot. Numerous environmental and other samples are taken for laboratory analysis, and the location and purpose of each is carefully recorded. Large numbers of photographs are taken, both general views of the progress of excavation and detailed shots showing the contexts of finds. Since excavation is a destructive process, this makes it imperative that the recording methods are both accurate and reliable. Unlike many other applications of information systems, it simply is not possible to go back and re-check at a later date [5]. Large quantities of archaeological data generated during the abovementioned two stages can be harvested by ArchDLs, organized, and stored to be available to researchers outside a project (site) – without substantial delay. After excavation, information stored in ArchDLs is analyzed, and helps archaeologists to test hypotheses. For example, if archaeologists retrieve records of corn artifacts from an ArchDL, they might hypothesize that the former residents were farmers, and test their hypothesis with soil sample data using statistical analysis tools provided by the ArchDL. This hypothesis is a scenario involving archaeologists, the historical community (farmers),

and finds (corn samples). Other hypotheses are scenarios describing relationships among historical communities. For example, if there are large collections of jars of the same style found in two nearby sites, archaeologists might hypothesize that people in these two sites (cities) used the jars to carry things in commercial trade. Thus, primary archaeological data, managed with powerful tools in ArchDLs, help archaeologists find physical relationships between excavation contexts, develop a structural history of a site, and extend the understanding of past material cultures and environments in the area. Data generated from the sites' interpretation then provides a basis for future work including publication, museum displays, and, in due course, input into future project planning.

Besides supporting archaeologists in their work as described above, ArchDLs provide services for the general public. A student interested in a Near Eastern site can access all the archaeological information about it by browsing or by using complex retrieval criteria that take account of both intrinsic attributes of items and their extrinsic spatial and temporal interrelationships. Further, she can view the information, e.g., organized in a spatial hierarchy or map that facilitates navigation among archaeological items at various spatial scales. She can click on items to show more detail; to display photographs, maps, diagrams, or textual documents; and to jump to other items.

3.   Spaces

One important spatial aspect of ArchDLs is the geographic distribution of found artifacts, which are located in four-dimensional spatial continua, the fourth one being the temporal (as inferred by the archaeologists). Metric or vector spaces are used to support retrieval operations, calculate distance, and constrain searches spatially. Another space-related aspect deals with user interfaces, or with 3D models of the past.

4.   Structures

Structures represent the way archaeological information is organized along several dimensions. Archaeological information is spatially organized, temporally sequenced, and highly variable. Examples include site organization, temporal order, and taxonomies of specific unearthed artifacts like bones and seeds. The structures of sites present simply and consistently the basic spatial containment relationship at every level of detail, from the broadest region of archaeological interest to the smallest aspect of an individual find. Generally, specific regions are subdivided into sites, normally administered and excavated by different groups. Each site is further subdivided into partitions, sub-partitions, and loci, the latter being the nucleus of the excavation. Materials or artifacts found in different loci are organized in containers for further reference and analysis. The locus is the elementary volume unit used for establishing archaeological relationships. Archaeological relationships between loci are from both the vertical and horizontal points of view. The first is given by reference to loci above and below a given locus, the second by coexisting loci (loci located at the same level). The archaeological relationship is related to the temporal succession of various events of construction, deposition, and destruction. Temporal sequencing of archaeological items involves linking items to form a stratigraphic diagram of the kind developed in the 1970s by Edward Harris (http://www.harrismatrix.com/) and now used by many archaeologists. A "Harris Matrix" is a compact diagram representing the essen-

tial stratigraphic relationships among all the items; it shows the chronological relationship between excavated layers and contexts. In general, if two layers are in contact with each other and one lies over the other, then the upper layer is chronologically later. This is the basis on which the structural history of a site is founded. The construction of this diagram and its subsequent use in the interpretation of structural phases is central to both the understanding of the site during excavation and to the post-excavation process [1]. Spatial and stratigraphic relationships among archaeological items can be regarded as extrinsic attributes (inter-item relationships) [6]; intrinsic attributes are those describing the items themselves. Finally, since archaeological information is highly variable, items observed in a typical excavation may fall into a wide variety of different classification systems, and may exhibit many idiosyncrasies.

5.   Streams

In the archaeological setting, streams represent the enormous amount of dynamic multimedia information generated in the processes of planning, excavating, analyzing, and publishing. Examples include photos and drawings of excavation sites, loci, or unearthed artifacts; audio and video recordings of excavation activities; textual reports; and 3D models used to reconstruct and visualize archaeological ruins.

4   A 5S-Based Metamodel for Archaeological Digital Libraries

With key requirements for ArchDLs summarized in the previous section, we can proceed to constructively define a minimal ArchDL metamodel. A domain-specific metamodel is a generic model which captures aspects specific to the domain at hand. We build upon the definition of a minimal DL as formally defined in [4] and extend it with concepts specific to the archaeology domain. Following our minimalist approach, we only define essential concepts without which we think a DL cannot be considered an ArchDL. The concepts and definitions are illustrated in Figure 2. Each concept is enclosed in a box labeled with the number of its formal definition (1-10 as below or starting with "A.", and in blue, if given in [4]). The main extensions have to do with the fact that: 1) most archaeological digital objects are surrogates of real-world artifacts; and 2) these artifacts are found within a social-temporal-spatial context.

Notation: Let $L=\bigcup D_k$ be a set of literals defined as the union of domains D_k of simple data types (e.g., strings, numbers, dates, etc.).
Let *SpaP1* be a tree with a vertex set $\{ v_i \mid i=1,2,...7\}$; an edge set $E_{SpaP1}=\{(v_i ,v_{i+1})\mid i=1,2,...6\}$; a labeling function F_{edge1}: $E_{SpaP1} \rightarrow L_{SpaP1}$, where $L_{SpaP1}=\{\text{'}contains\text{'}\}$; a labeling function F_{node1}: $\{ v_i \mid i=1,2,...7\} \rightarrow V_{SpaP1}$, where $V_{SpaP1}=\{F_{node1}(v_i) \mid i=1,2,...7\}$,
$F_{node1}(v_1)=$ '*region*', $F_{node1}(v_2)=$ '*site*', $F_{node1}(v_3)=$ '*partition*', $F_{node1}(v_4)=$ '*subpartition*', $F_{node1}(v_5)=$ '*locus*', $F_{node1}(v_6)=$ '*container*', and $F_{node7}(v_7)=$ '*find*'.
Let *SpaP2* be a set: $SpaP2=\{\text{'}above\text{'}, \text{'}below\text{'}, \text{'}coexisting\ with\text{'}\}$.

Let *TemP* be a tree with a vertex set $\{u_1, u_2\}$; an edge set $E_{TemP}=\{(u_1, u_2)\}$; a labeling function F_{edge2}: $\{u_1, u_2\} \rightarrow L_{TemP}$, where $L_{TemP}=\{$ *'detailed by'*$\}$; a labeling function F_{node2}: $\{u_i\} \rightarrow V_{TemP}$, where $V_{TemP}=\{$ *'period, 'chronology'*$\}$, $F_{node2}(u_1)=$*'period'*, and $F_{node2}(u_2)=$*'chronology'*.
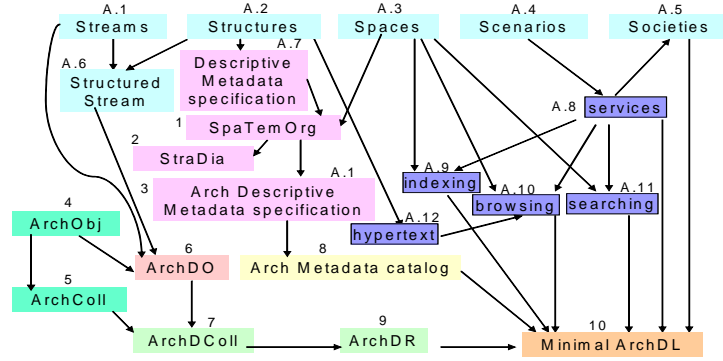

Fig. 2. Minimal archaeological DL in the 5S framework.

Definition 1: A ***Spatial Temporal Organization (SpaTemOrg)*** is a descriptive meta-data specification (see def. 12 in [4]), ***SpaTemOrg*** $= ((V, E), R \cup L, F)$, such that $\forall e = (u,v) \in E$, where $u,v \in V$, $F(u) \in R \cup L$, $F(v) \in R \cup L$, $F(e) \in V_{SpaP1} \cup V_{TemP} \cup SpaP2$.

Example 1.1: Given u, v_1, v_2 $\in V$, $F(u)=$ *'Bone1'*, $F(v_1)=$ *'Jordan Valley'*, $F(v_2)=$ *'Nimrin'*, $x=F((u, v_1))=$ *'region'*, $y=F((u, v_2))=$ *'site'*, $F_{SpaP1}((x, y))=$ *'contains'*, expression *('Bone1', (region: 'Jordan valley'), (site: 'Nimrin'))* means *'Bone1'* was excavated from the Jordan valley, which contains the Nimrin site (see Fig. 3).
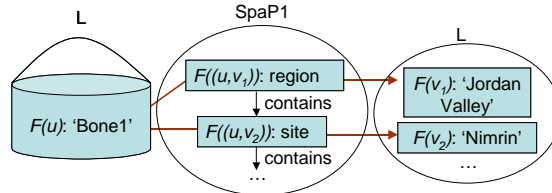

Fig. 3. Example 1.1 of *SpaTemOrg*

Example 1.2: Given u, v_1, v_2 $\in V$, $F(u)=$ *'Bone1'*, $F(v_1)=$ *'Middle Bronze'*, $F(v_2)=$ *'2000B.C. – 1500B.C.'*, $x=F((u, v_1))=$ *'period'*, $y=F((u, v_2))=$ *'chronology'*, $F_{TemP}((x, y))=$ *'detailed by'*, expression *('Bone1', (period: 'Middle Bronze'), (chronology: '2000B.C. – 1500B.C.'))* means *'Bone1'* was excavated from a deposit made in the Middle Bronze age, which has range 2000B.C. – 1500B.C. (see Fig. 4).
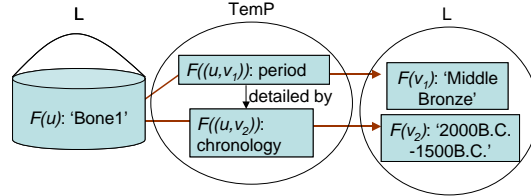
Fig. 4. Example 1.2 of *SpaTemOrg*

Definition 2: A ***Stratigraphic Diagram (StraDia)*** is a ***SpaTemOrg*** with a labeling function $F_{stra}: (V \cup E) \rightarrow L_{stra}$, where L_{stra} = *{'contemporary with', 'later than'}*, and two relations, $\cong$ and $\leq$, are defined on V:

   1) $\cong$ relation is:

     a) reflexive: $\forall v \in V : v \cong v$ b) symmetric: *if $v \cong u$, then $u \cong v$*

     c) transitive: *if $v \cong u$ and $u \cong w$, then $v \cong w$*

   2) $\leq$ relation is:

     a) reflexive: $\forall v \in V : v \leq v$

     b) anti-symmetric: *for all v, u $\in V$ if $v \leq u$ and $u \leq v$ then u=v*

     c) transitive: *if $v \leq u$ and $u \leq w$, then $v \leq w$*

such that $F(\cong)$ = *'contemporary with'*, $F(\leq)$ = *'later than'*, $\forall e = (u,v) \in E$, where $u, v \in V$, $F(u) \in R \cup L$, $F(v) \in R \cup L$, $F(e) \in L_{stra} \cup SpaP2 \cup V_{TemP}$.

Example 2.1: Given *v, u $\in V$, F(u)= 'locus1', and F(v)= 'locus2', $F_{stra}((u,v))$= $F(\leq)$ = 'later than', F((u,v))= 'above' $\in SpaP2$,* expression *('locus1' $\leq$ 'locus2')* means that *locus1* was later than *locus2*; and expression *('locus1', above, 'locus2'))* means that *locus1* was above *locus2* (see Fig. 5).
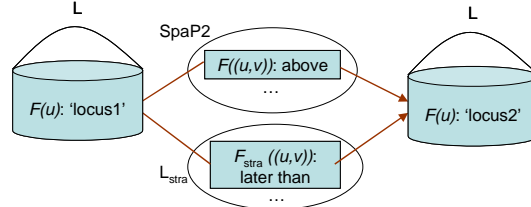

Fig. 5. Example 2.1 of *StraDia*

Definition 3: An ***Archaeology Descriptive Metadata specification: Arch_dm*** $\in \{SpaTemOrg\}$.

Definition 4: An ***Archaeology Object in real world (ArchObj)*** is a unit of observation generated by an archaeological activity (e.g., an archaeological town site, tomb, skeletal material, pottery, etc.).

Definition 5: An ***Archaeology Collection (ArchColl)*** is a tuple: *Arch-Coll=($h_{ArchColl}$, {ArchObjc})*, where $h_{ArchColl} \in H$, and *H* is a set of universally unique handles; *{ArchObjc}* is a set of archaeology objects in the real world.

Definition 6: An ***Archaeology Digital Object (ArchDO)*** is a tuple: *ArchDO=(h, SM, ST, StructuredStreams, SurrogateObj),* where

1) $h \in H$, where H is a set of universally unique handles;
2) $SM = \{sm_1, sm_2, \ldots, sm_n\}$ is a set of streams;
3) $ST = \{st_1, st_2, \ldots, st_m\}$ is a set of structural metadata specifications;
4) *StructuredStreams* $= \{stsm_1, stsm_2, \ldots, stsm_p\}$ is a set of *StructuredStream* functions defined from the streams in the *SM* set and from the structures in the *ST* set.
5) *SurrogateObj:* a function $h \rightarrow \{ArchObj_1, ArchObj_2, \ldots, ArchObj_k\}$ maps a handle h to an archaeology object in the real world, *SurrogateObj(h).*

Definition 7: An ***Archaeology Digital Collection (ArchDColl)*** is a tuple: *ArchDColl=($h_{ArchDColl}$, {ArchDO}, SurrogateColl),* where $h_{ArchDColl} \in H$, a H is a set of universally unique handles; *{ArchDO}* is a set of archaeology digital objects with handles in H. Let $Coll \in 2^{\{SurrogateObj(h)\}}$, where h is the handle of *ArchDO*; *SurrogateColl* is a function $h_{ArchDColl} \rightarrow Coll$ that maps handle $h_{ArchDColl}$ to a real world archaeology collection. Functions *SurrogateObj* and *SurrogateColl* are illustrated in Fig. 6.
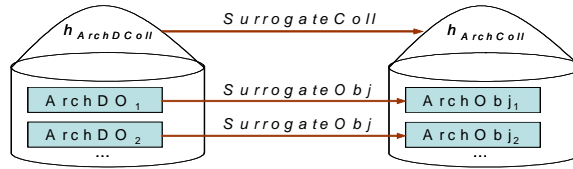

Fig. 6. Functions *SurrogateObj* and *SurrogateColl*

Definition 8: An ***Archaeology metadata catalog (ArchDM$_{ArchDColl}$)*** for an ArchDL collection *ArchDColl* is a set of pairs $\{(h, \{Arch\_dm_1, Arch\_dm_2, \ldots, Arch\_dm_i \})\}$, where $h \in H$ and each $Arch\_dm_i$ is an archaeology descriptive metadata specification.

Definition 9: Let *DColl* $=\{ArchDO_1, ArchDO_2, \ldots, ArchDO_k\}$ with k handles in H. An ***Archaeology Digital Repository (ArchDR)*** is a tuple: *ArchDR=(R, get, store, del),* where $R \subseteq 2^{DColl}$, and "get", "store", and "del" are functions over the repository (see def. 19 in [4] for details on these functions).

Definition 10: An ***Archaeological Digital Library (ArchDL)*** is a tuple: *ArchDL=(ArchDR, ArchDM, Serv, Soc),* where

1) *ArchDR* is an archaeology digital repository;
2) *ArchDM*$=\{ArchDM_{ArchDColl_1}, ArchDM_{ArchDColl_2}, \ldots, ArchDM_{ArchDColl_k}\}$ is a set of archaeology metadata catalogs for all archaeology digital collections *{ArchDColl_1, ArchDColl_2, \ldots ArchDColl_k}* in the repository;
3) *Serv* is a set of services containing at least indexing, searching, and browsing;
4) *Soc= (SM $\cup$ Ac, R),* where *SM* is a set of service managers responsible for running DL services, $Ac \subseteq \{Archeologist, GeneralPublic\}$ is a set of actors that use those services, and R is a set of relationships among *SM* and *Ac.*

5 Case Studies: Application of 5S to Archaeological DLs

In the last two sections, 5S was used to provide both an informal and a formal ArchDL model. In this section we use two archaeological information systems of ETANA projects (http://www.etana.org/) as case studies to 1) show the use of 5S as an analytical tool helpful in comprehending specific ArchDL phenomena; and 2) illustrate the use of 5S for requirements gathering and modeling in ArchDL development.

5.1 Virtual Nimrin

Tell Nimrin (TN) is an archaeological town site at Shuna South, Jordan, north of the Dead Sea, in the Jordan Valley. The digital presentation of TN, virtual Nimrin (VN) (http://www.cwru.edu/affil/nimrin/), at Case Western Reserve University, is supervised by director James W. Flanagan.

1. Societies

VN was designed for the general public as well as research specialists. Other communities addressed include: directors, core field staff (square supervisors, technical archaeologists, disciplinary specialists, assistant staff, and managers), and VN website developers/viewers from a score of museums, research institutions, colleges, and universities in Jordan and the United States.

2. Scenarios

Each of the communities involved in the VN society is engaged in various tasks and activities. Core field staffs were responsible for the actual work of excavation and recording. For example, in the field, unearthed bones were bagged separately, daily, with a feature and field specimen number which could be cross referenced with associated ceramics. These bagged bones were transported to field laboratories to be dry brushed, washed when necessary, and separated into generalized categories such as large, medium, or small mammals; fish and birds, etc. To advance and enhance digital recording, digital photography and additional programming were used. Project directors pursued geological and archaeological research by analyzing the field survey and excavated record, testing hypotheses, and publishing preliminary and final reports. For example, they found there was a reduction in percentages of bones of hogs over time at Tell Nimrin, and hypothesize that the reason probably was the introduction of religious taboos against eating pork. VN website developers built systems to allow users to interact with and interpret the site without being constrained by the director's view. General users may be interested in taking the tour of the VN website and in viewing museum quality artifacts and major finds, while specialists may want to interact with or download the databases created from the field records of excavation.

3. Spaces

TN's geographical setting marks the intersection of N-S and E-W arteries in the Jordan Valley approximately 12km north of the Dead Sea and 8km from the Jordan River due east of Jericho. It is located at $35^{o}37^{'}30^{''}$ east longitude and $31^{o}54^{'}00^{''}$ north latitude with a Palestinian grid reference of 2094E/1451N. The mound stands 12.75m

high on the south bank of the Wadi Nimrin. Its base is 200m below sea level. GPS was used in a geological survey, such as to document the regression of the ancient freshwater Lake Lisan that once filled the Jordan Valley, and to determine how the change from a freshwater to a saline body affected the pattern of ancient settlements in the region. Other space aspects of VN are TN's coordinate system (site grid and identification of squares) displayed in the topographical drawing, and VN's user interface.

4. Structures

Structures of VN include its relational database, TN's site organization, and TN's stratigraphic diagram, from which a temporal sequence was derived. Spatial and temporal description of records in the database is specified according to TN's polar point grid site organization and site chronological order. The 00/00 point was set at the highest elevation of the mound which was central to its N/S and E/W expanse as well. From there, the site was divided into quadrants, which were subdivided into $5m \times 5m$ squares, each labeled according to the point closest to 00/00. For instance, N40/W20 identified both the point 40m north and 20m west of 00/00 and the five meter square to the north and west of that point. Stratigraphical relationship analysis has identified eight major strata. They are: Modern (stratum VIII), Mamluk (stratum VII), Late Byzantine/Umayyad (stratum VI), Roman/Byzantine (stratum V), Persian (stratum IV), Iron II (stratum III), Iron I (stratum II), and Middle Bronze (stratum I). They clarified TN's long history as an agricultural town site and indicated TN was a substantial settlement, inhabited continuously for the past 4 millennia, except for a 500 year period.

5. Streams

VN deals with various streams, such as drawings and photos of (parts of) TN, publications of preliminary (final) reports, and tuples of primary data in the database.

Virtual Nimrin (VN) Formal Model

Virtual Nimrin is a tuple: VN = *(VN_R, VN_DM, VN_Serv, VN_Soc)*, where

1) *VN_R* is an archaeological digital repository having Tell Nimrin's digital collections of animal bones, seeds, etc. – *DCollBone, DCollSeed, ..., DCollObj*.

2) *VN_DM={VN_DM$_{DCollBone}$, VN_DM$_{DCollSeed}$, ..., VN_DM$_{DCollObj}$}* is a set of archaeology metadata catalogs for all archaeology digital collections in VN, where *VN_DM$_{DCollObj}$* is a metadata catalog for digital collection *DCollObj*. Let *VN_dm$_{DCollObj}$* be a descriptive metadata specification for digital objects in *DCollObj*. *VN_dm$_{DCollObj}$* ∈ *{SpaTemOrg}* ∪ *{VN_Dobj}*, where

   a) *{Jordan Valley', 'Nimrin', 'quadrant', 'square', 'locus', 'bag'}* ∪ *{'Ottoman-Modern', 'Islamic', 'Byzantine', 'Late Hellenistic-Roman', 'Persian', 'Iron II', 'Iron I', 'Middle Bronze'}* ∈ *L*. (See examples in Figures 3 and 4 for reference.)

   b) *VN_Dobj= ((V, E), R, F)* is an archaeological-object-specific descriptive metadata specification. If *DCollObj* is a digital collection of animal bones, then $\forall e = (u,v) \in E$, where $u \in V$, $v \in V$, *F(u)*∈*R*, *F(v)* ∈*L*, and *F(e)* ∈ *{'boneName', 'animalName'}* ∈ *L*

3) *VN_Serv={browsing, searching}*

4) *VN_Soc= ({VN_ServiceManager} ∪ {VN_Actor}, R) ∪ {HistoricNimrinResidence} ∪ {PrehistoricNimrinResidence}, R)*,
   where *{director, fieldStaff, specialist, student,…} ⊂ {VN_Actor}, {VN_BrowseManager, VN_SearchManager} ⊂ {VN_ServiceManager}, {browse=(student ✕ VN_BrowseManager, browsing), search=(specialist ✕ VN_SearchManager, searching), guide=(director✕fieldStaff, Φ)} ⊂ R*

5.2 Halif DigMaster

The site of Tell Halif, located in southern Israel, is the focus of the Lahav Research Project. Halif DigMaster (HD) is an online archaeological database that offers access to a collection of Persian/Classical (and some Iron II Age) figurines recovered in excavation from Tell Halif (TH).

1. Societies
HD was developed to disseminate archaeological information to the academy and to the public. Societies of HD include the communities who excavated the figurines from Tell Halif, provided HD with a preliminary presentation of the excavated material, collaborated with HD on resource sharing, or developed and/or employed HD.

2. Scenarios
Scenarios of HD can be those related to activities such early publication, collaborative publication, 3D publication, and other usage scenarios of HD. The overlong lag between discovery and publication is an embarrassment for archaeology of the ancient Near East. To address this embarrassment, staff of the Lahav Research Project made graphic forms of excavated objects available to the scholarly and professional world prior to final publication, while field work was still continuing. Scenarios of inviting excavators with non-public or incompletely published collections to add their materials to HD allow several excavation teams to share resources. Scenarios of using 3D technology such VRML (Virtual Reality Modeling Language) overcome the limitations of "flat" photographs on screen. Usage scenarios of HD describe services such as browsing and searching the HD database.

3. Spaces
Major spaces in HD are the geographical setting of Tell Halif (located at 34°52' east longitude and 31°23' north latitude, at coordinates 1373/0879 on the Palestinian grid) and HD's user interface. Another space is of the VRML models of artifacts.

4. Structures
Structures in HD include the relational database, Tell Halif's site organization, and TH's strata relationship. TH has been surveyed and plotted in relationship to the standard regional grid. Each of the major sections being worked is called a "field", which is further divided into a number of more or less standard 5m✕5m areas. TH consists of seventeen major occupation strata, one built atop another to a depth of more than six meters. Those strata revealed that TH has a history of occupation began in the

Chalcolithic era (3500B.C. – 3200B.C.) down to the modern settlement of Kibbutz Lahav (founded in 1963 A.D.).

5. Streams

Streams in HD are full-scale color photographs, ink drawings, QuickTime VR movies, plans and soil profiles, as well as summary reports for excavation units.

Halif DigMaster (HD) Formal Model

Halif DigMaster is a tuple: HD = *(HD_R, HD_DM, HD_Serv, HD_Soc),* where

1) *HD_R* is an archaeological digital repository having Tell Halif's digital collection of figurines, denoted as *DCollFig.*

2) *HD_DM={HD_DM$_{DCollFig}$}* is a set of archaeology metadata catalogs for ArchDL collections of figurines, where *HD_DM$_{DCollFig}$* is a metadata catalog for digital collection *DCollFig.* Let *HD_dm$_{DCollFig}$* be a descriptive metadata specification for digital objects in *DCollFig. HD_dm$_{DCollFig}$* ∈ *{SpaTemOrg},* where
   {'Southern Israel', 'Halif', 'field', 'area', 'locus', 'basket'} ∪ *{'Persian'}* ∈ *L.*

3) *HD_Serv={browsing, searching}*

4) *HD_Soc= ({HD_ServiceManager}* ∪ *{HD_Actor}* ∪ *{PersianHalif}, R),* where *{director, fieldStaff, specialist, student}* ⊂ *{HD_Actor}, {HD_BrowseManager, HD_SearchManager}* ⊂ *{HD_ServiceManager}.* We denote the community that made the Persian figurines excavated from Tell Halif as *FigMaker*, and denote the persons who those figurines represent (as surrogates) as *FigSurrogate.* Then *{FigMaker, FigSurrogate}* ⊂ *{PersianHalif}; {browse = (student* × *HD_BrowseManager, browsing), search=(director* × *HD_SearchManager, searching), describe=(specialist* × *FigSurrogate,* Φ *)}* ⊂ *R.*

6 Conclusion

DLs and archaeology have inherently interdisciplinary natures. This makes an ArchDL an even more complex information system and the task of formally defining it difficult. In this paper, we address this problem, defining a minimal ArchDL by applying and extending a DL formal framework – 5S. Our definition serves as the foundation for our enhanced ETANA-DL (http://feathers.dlib.vt.edu) prototype, now being refined so as to result from semi-automatic DL generation. Future work will include modeling distributed DLs, possibly including P2P approaches as in OCKHAM (http://www.ockham.org/), and developing assessment measurements for domain specific integrated DLs such as ArchDL.

References

1. Finkelstein, S., Ussishkin, D. and Halpern, B. Monograph Series of the Institute of Archae-ology, Tel Aviv University, 2000.
2. Gonçalves, M.A. Stream, Structure, Space, Scenarios, and Societies (5S): A Formal Digital Library Framework and Its Applications, PhD Diss., Virginia Tech Dept. Comp. Sci., 2004.
3. Gonçalves, M.A. and Fox, E.A., 5SL: a language for declarative specification and genera-tion of digital libraries. In Proc. JCDL 2002, 263-272.
4. Gonçalves, M.A., Fox, E.A., Watson, L.T. and Kipp, N.A. Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. ACM TOIS 22(2):270-312.
5. Ryan, D.N. Managing Complexity: Archaeological Information Systems Past, Present and Future. http://www.cs.kent.ac.uk/people/staff/nsr/arch/baas.html
6. Schloen, J.D. Archaeological Data Models and Web Publication Using XML. Computers and the Humanities, 35 (2). 123-152.