Dublin Institute of Technology
ARROW@DIT

Conference papers

Digital Media Centre

2008-02-01

Knowledge Management for Disparate Etruscan Cultural Heritage

John McAuley
Dublin Institute of Technology, john.mcauley@dit.ie

James Carswell
Dublin Institute of Technology, jcarswell@dit.ie

Follow this and additional works at: http://arrow.dit.ie/dmccon

Part of the Computer Sciences Commons, and the History of Art, Architecture, and Archaeology Commons

Recommended Citation
McAuley, J., Carswell, J.: Knowledge Management for Disparate Etruscan Cultural Heritage. icds, pp.70-74, Second International Conference on the Digital Society, 2008.

This Conference Paper is brought to you for free and open access by the Digital Media Centre at ARROW@DIT. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@DIT. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License

Dublin Institute of Technology
ARROW@DIT

Articles

Digital Media Centre

2008-02-01

Knowledge management for disparate Etruscan cultural heritage

John McAuley
Dublin Institute of Technology, john.mcauley@dit.ie

James D. Carswell
Dublin Institute of Technology, jcarswell@dit.ie

Recommended Citation

McAuley, John and Carswell, James D.:Knowledge management for disparate Etruscan cultural heritage. ICDS'08 Paper ID: 22-1228467117.

This Conference Paper is brought to you for free and open access by the Digital Media Centre at ARROW@DIT. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@DIT. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie.

Knowledge Management for Disparate Etruscan Cultural Heritage

John McAuley, Dr. James D. Carswell
Digital Media Centre, Dublin Institute of Technology,
{john.mcauley,jcarswell}@dit.ie

Abstract— This paper introduces the *TARCHNA* approach to managing and presenting contextualised heritage content. The system uses the CIDOC CRM ontology to consolidate a virtual repository of geographically disparate heritage databases and present a holistic view of a fragmented heritage. While previous approaches to presenting heritage collections have focused on the browse and search paradigm, the *TARCHNA* system uses narrative concepts as a means of presenting and re-using contextualised heritage artefacts within a broader cultural setting.

***Index Terms*—Information and Knowledge Management, Ontology, Narrative, Context**

I. Introduction

Cultural heritage artefacts, such as archaeological finds, are normally housed in disparate, often geographically remote museum collections. As such, the typical museum visitor, wishing to develop a deeper understanding of a heritage domain, is often frustrated by collections being fragmented across numerous heritage institutions. In addition, individual museum exhibition space is limited, resulting in artefacts lying in storage and away from the public eye for long periods of time. Increasingly however, museums and heritage institutions are investing resources in digitizing their collections. While much work has been carried out in the area of standards for digital cultural heritage (English Heritage, The Getty, ADS – Archaeological Data Standards), there is still no commonly agreed consensus on information and knowledge management for this "new" digital heritage information.

As with other areas of information management, curators and museum professionals use a variety of approaches and systems to manage their digitised content. The conventional problems that burden the interoperability of heterogeneous datasets are therefore highly significant to the domain of cultural heritage. Supplementary is the difficulty of "context": i.e. presenting heritage artefacts from a broader perspective and within their original context. Antecedent approaches have addressed this problem by focusing on presenting a united view of museum collections. Conversely, we suggest capturing a domain expert's knowledge by way of narrative presentations to convey a unified and contextualised portrayal of a cultural heritage domain. Although the *TARCHNA* project focuses on Etruscan heritage, the approach is considered general enough to be applied across the sphere of knowledge management for cultural heritage in general.

This paper gives a brief overview of the *TARCHNA* system architecture, the tiered components, and reasons behind the approach. An explanation detailing the issues of context is provided, plus the proposed solution of using ontologies to describe collections, cultural and narrative concepts. The paper concludes with a description of how mapping between the CRM ontology and a specific *TARCHNA* database instance was implemented.

T.Arc.H.N.A is a 3 year European project started in September 2004 and is partly funded by Culture 2000 (2004 -1488/001 -001 CLT-CA22)

II. *TARCHNA* System Architecture

Typical of enterprise information systems, the TARCHNA system is divided into a multi-tiered architecture whereby each tier supports a clear division of labour.

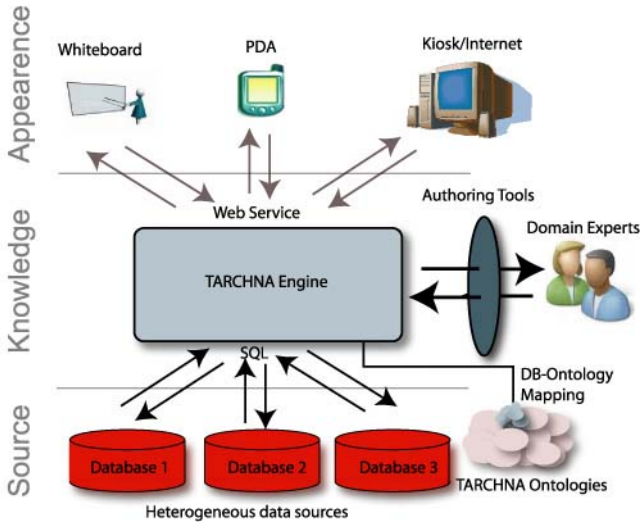

Fig 1. TARCHNA System Architecture

The *source tier* consists of several heterogeneous data-sources, each exhibiting a separate (i.e. dissimilar) database schema, and three Ontologies developed upon the CIDOC CRM data standard [1]. The first, or *TARCHNA*, domain

ontology provides a common reference model on which to integrate the different database schemas. This integration takes the form of a "Database to Ontology" mapping, whereby the elements defined by the database schema are expressed in terms of the ontology's concepts. Unlike other efforts, such as ARTISTE [2], the database mappings are held in the ontology. In this way, the mapping information can be accessed in the same way as instances of the ontology classes - thus removing the need for an external procedure to access mapping files or altering the individual database schemas. This approach will be discussed in more detail in later sections. Two other ontologies, "Sphere of Knowledge" and "Narrative Ontology", also sit in the source tier of the system. All three ontologies are represented in the RDF formalism and stored in a sesame RDF store [3].

The *knowledge tier* is the central constituent of the system. It consists of several authoring tools, which support domain experts to develop narrative content, and the *TARCHNA* engine, which guides interaction between the *TARCHNA* domain ontology and the individual databases. The *TARCHNA* engine processes requests expressed as ontological concepts, and converts them into separate SQL statements relevant to each dataset. This process occurs at two different stages: firstly when authors wishing to write a narrative search the system for relevant artefacts on which to base their narrative content; and secondly when a request is accepted from the *TARCHNA* web service, the engine retrieves all narrative content related to a specific artefact, or conversely, retrieves all artefacts related to a specific narrative. The *TARCHNA* engine and authoring tools were developed in the Java programming language with the Protégé Ontology API [4],[5].

The *appearance tier* acts as the disseminating component of the system, and distributes data (i.e. artefacts) enhanced narrative to several multimedia devices known as virtual wings (VW). The XML web service technology is used as means of interfacing the appearance layer with the virtual wings. It was felt that a service orientated and platform neutral architecture (SOA) supports a clear demarcation between the internal workings of a virtual wing and the overall data model of the system. In this way, new virtual wings may be added without a reliance on proprietary software or adjustments to the system architecture. Virtual wings are currently conceived as operating within three possible spheres:

- Firstly, as contextualised panoramic images. This innovative approach supports a comprehensive way of integrating conceptual models, such as the *TARCHNA* Domain & Narrative Ontologies, into panoramic images. The approach specifies semantic hotspots or trigger points whereby a visitor can query the image and receive information from the semantic model. It offers a new paradigm for accessing and interacting with semantically contextualised multimedia [6].
- Secondly, as handheld interactive tools. As both GPS[1] and PDA[2] technologies evolve into lightweight and economical location aware handheld devices, it has become increasingly possible to develop high bandwidth GPS applications for mobile devices. This VW is thought of as offering visitors a unique opportunity to explore ancient sites through real-time GPS based digital narrative, and thought of as similar to a personal guide.
- Thirdly, as a customised virtual museum. This option consists of narrative rich multimedia based applications operating within a museum space.

III. Adding Context through Narrative

The heritage domain is to be understood as consisting of expressions, some of which are tangible and others less so. Tangible heritage is embodied in physical objects and artefacts that give an anthropological significance to a society or people. As tangible heritage is considered both representative and metaphorical, its context however remains abstract and intangible [7]. This context is an amalgam of what Svensson calls knowledge systems or life ways, and relates to an artefact but is not intrinsically part of one [8]. In order to understand the significance of an artefact requires it to be presented within a broader context.

Narrative is proposed as a way of reconciling physical artefacts with their original intention or historic context and, in our case, presenting a holistic impression of Etruscan heritage. The aim is to support a team of domain experts (archaeologists, researchers, etc.) develop narrative presentations, which describe artefacts and their context within Etruscan society. In discussions with several archaeologists, the problem of assigning context to digital artefacts was raised. They suggested that in a cultural heritage setting, an artefact's context can be understood as a combination of its function and role within a specific society. From this perspective, artefacts are presented as references to physical objects from the underlying datasets, many of which are accompanied by multimedia illustrations, while their context is woven into the narrative text and buttressed with ontology concepts, representing both function and role, from the Sphere of Knowledge (ontology).

IV. *TARCHNA* Approach to Ontology

The *TARCHNA* system uses ontologies to define narrative concepts and represent the domain to which they relate. Several distinctions were made to help formalise this process. Firstly, the domain was divided between aspects of tangible heritage, in the form of physical artefacts and monuments, and the broader concepts of Etruscan culture, such as economy, history, and religion. Each was represented by a separate ontology developed upon the CIDOC CRM data standard. The first, eponymously named the *TARCHNA* Domain

[1] Global Positioning System or GPS is a satellite navigation system.

[2] Personal Digital Assistants or PDA's are versatile handheld personal computers.

Ontology, is a formal definition of Etruscan artefacts and monuments. It functions as an umbrella ontology for the addition of supplementary datasets without the need for replication across repositories, while supporting a faceted search paradigm, and presenting the user with a unified view of a fragmented heritage. The ontology was developed in coordination with a team of archaeologists who have extensive experience of Etruscan antiquity.

The second ontology however is a less formal representation, and describes the broader concepts of Etruscan culture. The 'Sphere of Knowledge' Ontology exhibits weaker semantics by way of hierarchically ordered terms. An explanation of each is provided with a (natural language) scope note. The motivation in using a less formal approach lay with supporting a community of domain experts. It was felt that the community should be involved in, as much as possible, the initial development and continuous refinement of the ontology. In this way the community's knowledge may evolve, and consequently be reflected in the ontology, with the addition of new collections and narrative content. This method was successfully demonstrated by Srinivasan during his work on the Village Voice project where he approached the development of structured knowledge in terms of community participation and mutability [9]. He refers to the concept as fluid ontologies, or 'flexible knowledge structures that evolve and adapt to a communities' interest [10].

The third and final representation is the *TARCHNA* Narrative Ontology. This draws on much of the work by Mulholland and others when formally describing narrative concepts [11]-[13]. Narrative is thought of as an epistemological container for communicating heritage content. It does this by specifying several properties which tie together concepts from both the *TARCHNA* Domain Ontology and the Sphere of Knowledge in a single narrative presentation. The former describes artefacts by way of direct relations or characteristics of artefacts through indirect relations, while the latter discusses broader domain concepts which often represent the function and role of an artefact. *TARCHNA* Narrative is stored as class instances in the narrative ontology. In this way it is abstracted from the underlying datasets, but can still reference database objects via *direct* and *indirect* relations.

V. Mapping the CRM with Heterogeneous Data

The following section describes the approach taken to mapping the CRM ontology to a *TARCHNA* database developed in line with the project. The implementation was carried out by the project partners at the University of Milan. As mentioned previously, the CRM is an extensive property based (81 classes & 132 unique properties) cultural heritage, upper domain ontology. In 2006 it became an official standard (ISO 21127:2006) in promoting interchange amongst heterogeneous cultural heritage datasets. The intention of the CRM developers is, therefore, not to define the terminology of specific cultural heritage datasets but rather to facilitate information interchange and mediation.

A review of the literature in the area of ontology-database mapping illustrates several approaches adopted by projects similar to *TARCHNA*. The EPOCH network (http://www.epoch-net.org/), for instance, has developed a tool, called Archive Mapper for Archaeology or AMA, for the semi-automated mapping of archaeological archives to the CIDOC CRM. Similarly D2R MAP (http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2rmap/D2Rmap.htm) uses an XML-based language to describe mappings between relational databases and RDF based systems. While the museumFinland portal uses an XML schema notation to associate content from different content providers [14]. However, it was generally agreed that none of the above approaches suited that of the *TARCHNA* system. As indicated previously, it was felt that to reduce overhead when querying, it would be of more benefit to directly store mapping information as textual metadata (class instances) in the ontology itself.

There is an inherent complexity associated with archaeological databases. Frequently, as illustrated with the CRM, capturing information to generate knowledge engenders intricate conceptual structures, as archaeologists strive to capture an artefact's physical characteristics together with its environment and relationship with other findings. In discussions with archaeologists having extensive experience of Etruscan excavation, both monuments and findings were highlighted as the most salient features as regards to the tangible heritage of Etruscan culture. Therefore, the approach taken was to, in as much as way as possible, map these database entities directly to concepts from the CRM ontology.

The mapping information is encapsulated in two ontology classes, created solely for the purpose of mapping and neither is included in the original CIDOC specification, they are *DB_Class_Mapping* and *DB_Property_Mapping*. The first class describes the mapping information between the CRM classes and the database entities, while the second describes the CRM properties with database entity attributes. The *TARCHNA* engine uses this information for two purposes: firstly, the information is extracted during the authoring process to compile a set of XML files representing the mapping between the CRM and specific database entities (Fig. 2 illustrates a snippet of one such file). These files are used to narrow the author's query, which is then executed against the database. The results are correlated and returned to the author who can then forward to writing their narrative text. Secondly, the information is used when the end-user queries the system for either archaeological or narrative content. If, for instance, a user queries the system for musical instruments, the mapping information is used to retrieve all artefacts of type musical instrument and all narrative content related to those artefacts. Here we will discuss the first approach, which is implemented during the authoring process.

Firstly we briefly discuss the mapping process. The class *E22.Man-Made_Object* was mapped to the entity *Findings* in the *TARCHNA* database, while similarly the class *E24.Physical_Man-Made_Stuff* was mapped to the

Monuments entity in the *TARCHNA* database. Fig. 2 depicts a snippet of the Finding.xml file, used to represent the mapping information between the CRM class *E22.Man-Made_Object* and the Findings entity in the database. As can be seen a Finding or artefact is represented by a parent node of type *DB_Class_Mapping* and the related attributes as child nodes of type *DB_Property_Mapping*.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <DB_Class_Mapping tarchna_class_name="Finding" cidoc_class_name="E22.Man-
  Made_Object" stored_in="database" table="Findings" datatype="" SelectField="no"
  WhereField="">
  - <DB_Property_Mapping tarchna_property_name="finding provenance"
    cidoc_property_name="p53f.has_former_or_current_location"
    domain="Finding" range="Provenance"
    external_mapping="Findings.ProvenanceID = AreaProvenance.ProvenanceID"
    internal_mapping="" target_property="">
    - <DB_Class_Mapping entity="Finding" tarchna_class_name="Provenance"
      cidoc_class_name="E53.Place" stored_in="database"
      table="AreaProvenance" datatype="" SelectField="no" WhereField="">
      - <DB_Property_Mapping tarchna_property_name="provenance"
        cidoc_property_name="p2f.has_type" domain="Provenance"
        range="ProvenanceType" external_mapping=""
        internal_mapping="AreaProvenance.Provenance" target_property="">
        <DB_Class_Mapping tarchna_class_name="ProvenanceType"
          cidoc_class_name="E55.Type" stored_in="ontology"
          table="AreaProvenance" datatype="string" SelectField="no"
          WhereField="" />
        </DB_Property_Mapping>
      </DB_Class_Mapping>
    </DB_Property_Mapping>
```

Fig 2. An example of Findings.xml generated through the TARCHNA system which extracts textual mapping information from the ontology. Each entity is represented by a parent node and the related attributes as child nodes.

The next step was to map each entity's attributes with the most appropriate CRM properties. As illustrated in Fig. 2, the *TARCHNA* property finding class, describing the type of finding (e.g. votive offering, red figure, musical instrument), is mapped directly to the CRM property *p2f.has_type* through the internal_mapping tag. This tag is used to describe non-relational attributes, as in this instance the tag contains the value of *Findings.ClassID*. However, some attributes indicate relationships with other database entities, such as the property mapping finding provenance, while being mapped directly to the CRM property *p53f.has_former_or_current_location*, is described in the mapping information by the tag *external_mapping*, and is assigned *Findings.ProvenanceID = AreaProvenance.ProvenanceID*. The external_mapping tag indicates a relationship between database entities, as in this case the Findings entity is related to *AreaProvenance* via the relationship *Findings.ProvenanceID= AreaProvenance.ProvenanceID*. As a result, the *TARCHNA* engine can manipulate the mapping information to retrieve the appropriate database resources.

The system then uses several XSL transformations to generate faceted search interfaces (Fig. 3) for each of the database entities. The engine can now interact with the XML files without querying the ontology.

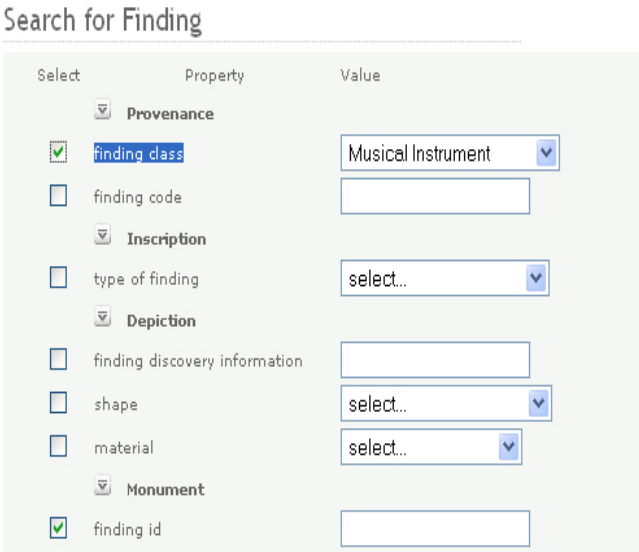

Fig 3. The findings faceted search interface, created through an XSL transformation using findings.xml.

Each facet that the author decides upon will result in the selectField tag of that property/attribute indicating a yes value (Fig. 4). Similarly, the wherefield tag will indicate the selected value chosen by the author. In Fig. 3 the author chose the *TARCHNA* property Finding Class and specifies search term as Musical Instrument. The resulting XML file show in Fig. 4 illustrates the author's query. The *TARCHNA* engine takes this file as an input and executes the query against the database, correlates the results and sends a single resource back to the author, who proceeds to writing the narrative text.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <DB_Class_Mapping tarchna_class_name="Finding" cidoc_class_name="E22.Man-
  Made_Object" stored_in="database" table="Findings" datatype="" SelectField="no"
  WhereField="" label="">
  - <DB_Property_Mapping tarchna_property_name="finding class"
    cidoc_property_name="p2f.has_type" domain="Finding" range="FindingClass"
    external_mapping="" internal_mapping="Findings.ClassID" target_property=""
    label="">
    <DB_Class_Mapping tarchna_class_name="FindingClass"
      cidoc_class_name="E55.Type" stored_in="ontology" table="Findings"
      datatype="string" SelectField="yes" WhereField="Musical Instrument"
      label="" />
    </DB_Property_Mapping>
  + <DB_Property_Mapping tarchna_property_name="finding id"
    cidoc_property_name="P47f.is_identified_by" domain="Finding"
    range="FindingID" external_mapping="" internal_mapping="Findings.FindingID"
    target_property="" label="">
  </DB_Class_Mapping>
```

Fig 4. XML file illustrating the refined query

VI. Conclusion

The *TARCHNA* project presented in this paper proposes a novel way of contextualising heterogeneous datasets through the construction and presentation of knowledge intensive narrative. The system hinges on an open approach to

information usability by promoting a clear separation of source, knowledge, and appearance. The multi-tiered architecture, while supporting semantic integration of heterogeneous datasets and avoiding data replication, provides a platform independent way to interact with and disseminate knowledge based narrative.

Currently, the system is being used by a number of European based archaeologists developing a suite of narrative discussing varying aspects of Etruscan heritage. While the approach was developed to support cultural institutions to amalgamate artefacts and present a holistic understanding of a specific heritage, it is not proprietary to subject matter or domain. The multi-tiered architecture supports data integration at both the procurement and dissemination stages, while the knowledge layer exploits narrative as a unifying platform, and presents both knowledge and data in an engaging format.

ACKNOWLEDGMENT

Seven project partners are involved in *TARCHNA*: the University of Milan in Italy, the Computer Science Department at Claude Bernard University in Lyon, France, the Institute of Archaeology/ Art Collections at University of Bochum, Ruhr, Germany, the Dublin Institute of Technology and University College Dublin both based in Ireland, the Department of Classical Archaeology at the University of Warsaw, Poland and the Aristotle University of Thessaloniki situated in Greece. Further information is available from the project website [15].

REFERENCES

[1] Crofts, N., et al., Definition of the CIDOC Conceptual Reference Model. 2005, ICOM/CIDOC. p. 94.

[2] Allen, P., R. Vaccaro, and G. Presutti, ARTISTE: An integrated Art Analysis and Navigation Environment. Cultivate: interactive, 2000(1).

[3] Broekstra, J., A. Kampman, and F.v. Harmelen. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. in International Semantic Web Conference. 2002. Sardinia, Italy.

[4] Gennari, J.H., et al., The evolution of Protege: an environment for knowledge-based systems development. Int. J. Hum.-Comput. Stud., 2003. 58(1): p. 89-123.

[5] Noy, N.F., R.W. Fergerson, and M.A. Musen, The Knowledge Model of Protege-2000: Combining Interoperability and Flexibility, in Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management. 2000, Springer-Verlag.

[6] Mazzoleni, P., et al. Towards a contextualized access to the cultural heritage world using 360 Panoramic Images. in Software Engineering and Knowledge Engineering 2006. San Francisco.

[7] A.K.Das. Intangible Heritage: Dynamics of Conservation in Museum of Living History. in ICTOP - International Committee for the Training of Personnel - Annual Conference (ICOM). 2004. Seoul, Korea.

[8] Svensson, T.G. Knowledge and context - the social life of objects in Museums and Intangible Heritage ICOM general conference. 2004. Seoul, Korea.

[9] Srinivasan, R. Village Voice: An Information-based Architecture for Community-centered Exhibits. in Museums and the Web. 2003. Toronto Ontario.

[10] Srinivasan, R. and J. Huang, Fluid ontologies for digital museums. Int. J. Digit. Libr., 2005. 5(3): p. 193-204.

[11] Mulholland, P. and T. Collins. Using Digital Narratives to Support the Collaborative Learning and Exploration of CH. in IEEE International workshop on Presenting and Exploring Heritage on the Web (PEH'02) in conjunction with the 13th International Conference and Workshop on Database and Expert Systems Applications (DEXA 2002). 2002. Aix-En-Provence, France. .

[12] Mulholland, P., T. Collins, and Z. Zdrahal. Story Fountain: Intelligent support for story research and exploration. in Intelligent User Interfaces (IUI'2004). 2004. Madeira, Portugal.

[13] Mulholland, P. and Z. Zdrahal. Knowledge Support for Story Construction, Exploration and Personalisation in Cultural Heritage Forums. in 14th International Conference and Workshop on Database and Expert Systems Applications (DEXA 2003). 2003. Prague, Czech Republic.

[14] Hyvonen, E., S. Saarela, et al. (2004). A Cultural Community Portal for Publishing Museum Collections on the Semantic Web. 16th European Conference on Artificial Intelligence (ECAI2004), Workshop on Application of Semantic Web Technologies to Web Communities, Valencia, Spain., CEUR Workshop Proceedings.

[15] TARCHNA. T.Arc.H.N.A. Towards Archaeological Heritage New Accessibility. 2005 [cited; TARCHNA project website]. Available from: http://www.tarchna.org/.