Directory of Hellenic Academic and Research Entities (D.H.A.R.E.)

Dimitris A. Dervos
Information Technology Dept.
Alexander Technology
Educational Institute (ATEI) of
Thessaloniki
P.O. BOX 141, 57400 Sindos,
Greece
dad@it.teithe.gr

Georgios Evangelidis
Dept. of Applied Informatics
University of Macedonia
P.O. BOX 1591, 54006
Thessaloniki, Greece
gevan@uom.gr

Kostas Ntonas
International Hellenic University
14th km Thessaloniki - Moudania,
57001 Thermi, Greece.
kntonas@gmail.com

Thanos Petridis
Aristotle University of
Thessaloniki
Library and Information Centre
AUTh Campus, 54124
Thessaloniki, Greece
thanos.petridis@gmail.com

Ilias Nitsos
Department of Library Science
and Information Systems
Alexander Technology
Educational Institute (ATEI) of
Thessaloniki
P.O. BOX 141, 57400 Sindos,
Greece
i.nitsos@libd.teithe.gr

ABSTRACT

The Directory of Hellenic Academic and Research Entities (DHARE for short) comprises an act undertaken under the Hellenic Libraries Link (HEAL-Link) project titled "Advanced Services for Open Access Digital Libraries" [1,2]. The aim is to design and develop a national system for the unique identification of research and academic entities (individual researchers, academic institutions, research centers, research organizations, etc.). The DHARE directory will be primarily used by services that incorporate in their functionality the unique identification of academic and research entities in Greece. Services like those of the institutional and/or subject repositories [5], current research information systems [6], research bibliography systems [7], etc.

The co-authors of the present work comprise the DHARE project support team and they herewith report on the main requirements specifications, system architecture, operational characteristics, and the plan to have DHARE come into service in late 2015.

Categories and Subject Descriptors

D.2.1 Software Requirements/Specifications, D.2.11 Software Architecture, H.3.5 Online Information Service

General Terms

Design, Standardization

Keywords

National Registry, Unique Identification Service, Academic and Research Entities

1. INTRODUCTION

In today's services interconnection and interoperability reality, the need to uniquely identify entities across Internet-based applications plays a decisive role for the operational reliability and the effectiveness of the latter. The DHARE project's aim is to design and develop a national directory for the unique identification of entities that comprise the academic and research community in Greece, namely: individual researchers, higher education academic establishments, research organizations/institutes, etc. By doing so, DHARE as a service is to be used to effectively resolve cases of typos, synonymy, and name abbreviations in third party services that conduct searches based on author/researcher names and/or names of academic/research organizations.

Focusing on just the unique identification of individuals (i.e. scientists, writers, artists, etc.), the relevant map of International Repositories Infrastructure wiki [8] lists a large number of projects undertaken at the national and/or corporate/institutional level, worldwide. One major development has been the appearance of the ORCID registry in the international scene [9]. ORCID first became operational in October 2012. It is an open, non-profit community-based effort enjoying the support of a large number of members including corporate bodies, publishers, universities, research institutes, etc. The service uniquely identifies researchers and relates them to published works they have (co-) authored; works that are registered with the major international scientific bibliography databases. Utilizing the ORCID persistent identifier the service distinguishes each one researcher from the rest, making possible the worldwide integration of its identification mechanism services in administering research workflows, manuscript submissions, grant applications, etc. The major development that ORCID represents for the international research community dictates the necessity to justify the need for developing DHARE as a national directory/registry service in Greece. This will be justified in the sequel, following the outline of the main operational characteristics of the DHARE service.

2. THE NEED

There is a clear need for a service that is to act as a reference directory to third party services, separating the complexity of the unique identification of academic and research entities pertaining to the operational logic of the latter. The need is best demonstrated by the following real life example:

The Hellenic Academic Libraries Link (HEAL-Link, [10]), through a license agreement with the Elsevier SCOPUS bibliography database, today has access to some 21,000 international research publication titles of various publishers [11]. The content of the SCOPUS bibliography database is dated from December 1998. A subset of the HEAL-Link SCOPUS (XML) content was used by the DHARE project support team to extract data on Greek researchers and Greek academic and research organizations listed in research works published during the 2000-2009 period. Table 1 outlines the profile of the data corpus considered.

Upon inspection, the processing of the HEAL-Link SCOPUS data has revealed all possible types of anomalies relating to the absence of an operational logic that would guarantee the unique identification of each one of the research entities involved. For example:

- More than one SCOPUS Author IDs are found to have been assigned to the same individual/researcher,

- More than one individuals/researchers are found to have been assigned the same SCOPUS Author ID

- More than one SCOPUS Affiliation IDs are found to have been assigned to the same research organization
- The same SCOPUS Affiliation IDs found to have been assigned to different research organizations
- Many individuals/researchers appear to be using the same email address, etc.

Indicative of the above is the fact that when searching for "Aristotle University of Thessaloniki", the "%aristotle%" criterion produced a result involving thirty seven (37) discrete SCOPUS affiliation ID values, whereas the "%aristotelian%" search criterion produced a result with six (6) discrete SCOPUS affiliation IDs. An analogous search for the University of Athens using the "%univ%of athens%" criterion produced a result involving forty three (43) discrete SCOPUS affiliation ID values.

Table 1: The HEAL-Link Elsevier SCOPUS data considered

Publ. Year	Records	Distinct Research Works
2000	18269	5922
2001	21030	6283
2002	23132	6820
2003	25240	7806
2004	29837	9125
2005	34936	10539
2006	42598	12528
2007	44845	13310
2008	47745	14211
2009	49947	15018
Totals:	**337579**	**101562**

Distinct Author IDs	Distinct Affiliation Ids
74122	5114

Figure 1 is indicative of the inherent ambiguity and (the consequent) lack of reliability of the SCOPUS database content when the latter is used to identify individuals/researchers.


Figure 1. **Ambiguity affecting the identification of Greek researchers in the HEAL-Link SCOPUS database**

The aforementioned absence of a sound support mechanism that guarantees uniqueness in the identification of academic and research entities should not be mistaken to only relate to the Elsevier SCOPUS bibliographic data content. During the past few years, a big effort has been undertaken by HEAL-Link and the central library of the Aristotle University of Thessaloniki to manually process bibliography data content from the Thomson-Reuters Web of Science database (WoS, [12]); The effort has been undertaken in order to make possible the unique identification of own (institutional) entities in the latter. The aim has been to improve on the reliability of the stated data in order for the latter to be incorporated into the Thomson-Reuters InCites service [13]. Starting from an initial number of 14962 distinct WoS affiliation IDs relating to the Aristotle University of Thessaloniki, 250 discrete affiliation IDs were established. With regard to the identification of the University academic/research staff, an initial set of 58000 WoS author IDs was reduced to correspond to a population of just 16700 distinct individuals. Equivalently, the typical Aristotle University researcher was found to have been assigned 3.47 WoS author IDs, on the average. On the basis of the Thomson-Reuters InCites experience, the Aristotle University has now established an in-house service that identifies its academic and research staff, utilizing a unique identifier named "University Registration Number".

The above justify (and lead to) the need for the establishment of a national directory service that is to uniquely identify academic and research entities not only within but also across relevant platform applications/services. A service that will make possible to guarantee, for example, that researcher-A who is registered to have published work-X in institutional repository-1 is the same individual as researcher-B who is registered to have published work-Y in institutional repository-2.

3.   DHARE ARCHITECTURE
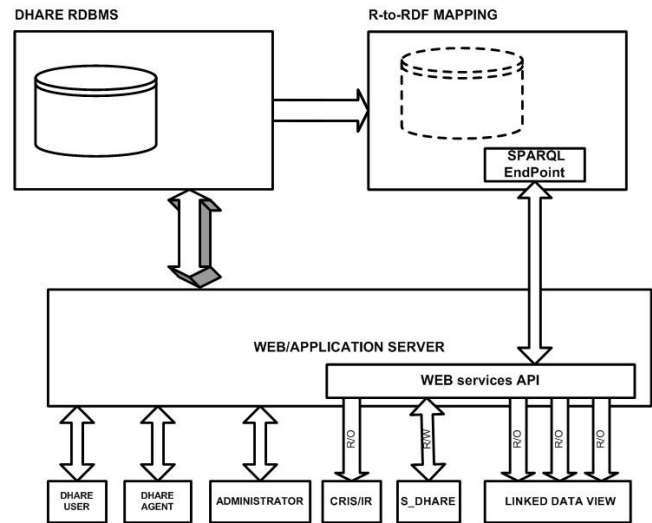
Figure 2. DHARE system architecture

The DHARE system architecture is outlined in Figure 2. It involves a typical database and web application server platform, plus an R-to-RDF mapping component via which supports a read-only linked data type view to a well-defined subset of the DHARE RDBMS data content. The web services application programming interface (API) supports read-only and read-write access to the database's data content. The former is to serve third party services on the Internet like subject and/or institutional repositories [5], CRIS systems [6], research bibliography systems [7], etc. In the beginning, the read-write API interface to the database's data content is to be used only by a satellite DHARE

(S_DHARE) for the purpose of updating the DHARE data content with data from international bibliography databases like the Elsevier SCOPUS [11], and the Thomson-Reuters Web of Science database (WoS, [12]), on the assumption that the corresponding license agreements are made with the corresponding organizations.

The DHARE project is currently (June 2015) at the stage where an international tender procedure is to determine the contractor who will design and develop the DHARE service shown in Figure 1. In this respect, it is still an open issue whether the R-to-RDF mapping sub-system shown in Figure 1 is to be implemented by means of a mapping file/mechanism, or whether it is to involve a (persistent) RDF repository. Both solutions involve their pros and cons; the R-to-RDF mapping sub-system comprising a pilot implementation for this initial version of the DHARE service, it is left with the contractor-to-be to decide, in collaboration with the DHARE project support team, the way it is to be implemented, using open source software. Once implemented and used in practice, the corresponding usage and performance measurements will be considered in the shaping of the next generation of the DHARE service.

Third party applications are expected to utilize the DHARE web services API subsystem in order to search for and retrieve unique identifiers for individuals and organizations. A characteristic example could be the use of a jQuery autocomplete widget in a submission form of a DSpace repository. The form will enable its user to conduct on-the-fly searches for DHARE IDs, via the web services API. The latter will return the results in JSON and/or XML data formats. It will be up to the client (i.e the third party service) to transform the results and present them to its user in suitable form, e.g. as a drop-down list.

4. USER ROLES AND TYPICAL OPERATIONAL INSTANCES

As seen in Figure 2, the DHARE system web interface is to support three main types of authorized access: administration, agent, and user access. System administrators will be HEAL-Link personnel who will be authorized to operate, configure, maintain, and update the service. DHARE agents will primarily be library personnel who will be authorized to conduct privileged operations like system users support tasks, plus batch and massive data content updating operations as (e.g.) in the case of having two departments in their organization merge into one, a task that involves the updating of registered data reflecting professional affiliations of the system users affected by the structural change in question. DHARE agents may be assigned additional privileges enabling them o maintain their organization's DHARE profile. The latter privilege may also come in separate, namely have DHARE users whose sole responsibility will be to maintain the profile of their organization, without being able to conduct the DHARE agent tasks outlined above. Last but not least, comes the primary target population of users that DHARE is to serve: individual academics, researchers, and authors of scientific works, in general. Users who will be able to create and updated their own DHARE profile, and conduct user profile merging operations when they realize that they have been assigned more than one DHARE IDs (i.e. user accounts with the system). In addition to the above, there will be a special type of (non-human, generic) API user. One that is to be used by third party applications and services to interact with DHARE, retrieve (via the R/O interface), or exchange data (via the R/W interface), and ultimately utilize DHARE IDs in order to uniquely identify academic and research entities within its own operational environment.

As stated in the previous, there are two primary academic and research entities to the DHARE data model: individual academics, researchers (E_I for short), and organizations (academic institutions, research centers/institutes, organizations: E_O for short).
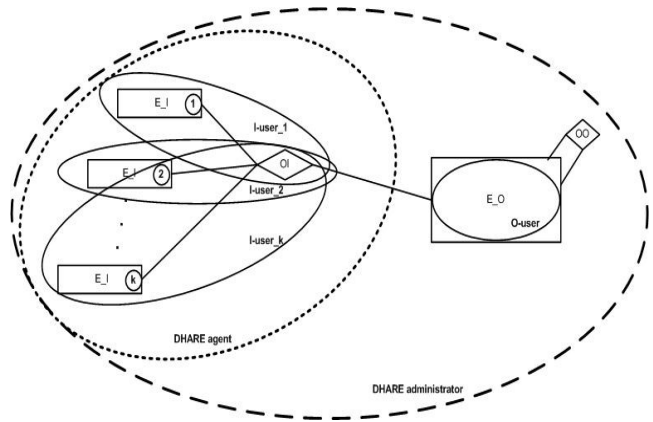

Figure 3. Primary DHARE entities, relationships and user roles

Figure 3 is indicative of the primary entities, their inter-relationships, and the user roles involved in the DHARE "topology" of tasks and data content administration procedures. Each one user (I-user in Figure 3) will be able to create and maintain his own E_I profile with the service. E_I user profile maintenance implies also the ability to maintain a list of professional affiliations (OI relationship in Figure 3). At this point, it is worth mentioning that the typical I-user will not be restricted to register professional affiliations only with system registered (i.e. DHARE coded) E_O entities. Equivalently, DHARE will not prevent, say, a high school teacher who conducts research and publishes his work in research journals from creating and maintaining his own E_I entity and, by doing so, establish/obtain his personal DHARE identifier (DHARE ID).

As it shown in Figure 3, DHARE agents will be authorized to create and access E_I user profiles, plus to update the data content of the latter (professional affiliations included). It is in this respect that special attention has been given to have the system specifications include the automatic maintenance of a detailed history (log) of update actions targeting each one data item of the E_I and E_O entities, as well as the inter-relationships involved. Each one I-user, for example, will be able to check the most recent update actions made to each one data item in his E_I profile, in a "who has updated what and in what way" manner.

The DHARE administrator(s) will be in charge of the E_O entities and their (horizontal) inter-relationships (OO in Figure 3), plus determine the DHARE users (DHARE agents included) who will be authorized to update E_O profiles: one O-user role assigned to each one E_O entity.

To roughly sketch the operational characteristics of the DHARE service, it helps to herewith enumerate its do's and don'ts:

- It will be acting as an one-stop service both to E_I users as well as to third part services: the E_I user profile will register identifiers used by third party services to identify the individual user in question (e.g. SCOPUS Author ID, WoS ResearcherID, ORCID, Aristotle University Registration Number, CRIS ID, etc.). By having a number of third party identifiers be assigned to the same DHARE ID, the service is

to facilitate the identification of each one of its E_I entity owners across a number of third party services.

• Operating the way it is designed to operate, the DHARE service will effectively relieve the third party services from the overhead of having to incorporate the complexity of unique identification of E_I and E_O entity types in their own business logic.

• With the use of DHARE ID, third party services will instantly have access to the given entity's name variations, professional affiliations, record of changes the entity in question has undergone (e.g. merges or splits, in the case of an organization, etc.)

• DHARE will not be the place for E_I owners to register own biography data like patents, participation in projects, research publications, etc.

• It will be at the discretion of each one E_I owner to specify the data items in his E_I profile that are meant to become publicly available.

• E_I-E_I, or E_O-E_O merges will not be 'broadcasted' /propagated to third party applications. The latter will continue to refer to the entities involved by means of the older (now obsolete) DHARE ID they register. However, when they do so, the information will be there (i.e. in the corresponding entity profile data content), informing of the change that has taken place.

• In the DHARE data model, E_O entities inter-relate to one another in horizontal fashion. There will not be structural hierarchies inherent to the typical E_O entity. For a example, in the case of a University E_O, Schools and Departments will be registered in the form of (hierarchical) sub-divisions; they will be registered in the form of non-hierarchically dependent, multi-valued organizational units.

5. PROJECT SUPPORT TEAM PRIORITIES AND TASKS

The role of the DHARE project support team is to pursue tasks that, among others, will facilitate the work to be undertaken by the project contractor. The latter will be the one to design and develop the DHARE service. For the final product to meet its user needs, the project support team have undertaken and they are currently carrying out the following tasks:

• ORCID API interface testing and prototyping. Developed a client script that queries the ORCID API and parses/processes the results of latter. The prototype allows for future extensions in the direction of importing ORCID originating data into the DHARE service.

• DHARE API prototyping. Utilized the experience of the tests conducted in relation with the ORCID API interface to prototype the DHARE API interface. Developed code to enable a typical institutional repository platform (DSpace) to retrieve data via a prototype DHARE service over the Internet.

• Established contact and close collaboration with the HEAL-Link national digital repositories support team based at the National Technical University of Athens (NTUA). Introduced DHARE identifiers into the HEAL-DSpace, a DSpace variant developed by NTUA. The latter is to be used by a large number of higher education universities and institutions in the

context of the national digital convergence operational program [3], under the co-ordination of the stated HEAL-Link, NTUA-based, support team.

• Established a DHARE users working group (UWG) involving representatives from five (5) academic libraries in Greece, namely: the libraries of (a) the Aristotle University of Thessaloniki, (b) the National Technical University of Athens, (c) the Alexander TEI of Thessaloniki, (d) the University of Macedonia and (e) the University of Patras. The DHARE project support team members tele-conference regularly with the DHARE UWG members, over the Internet.

Along the lines of the aforementioned set priorities and tasks undertaken, the DHARE project support team have developed a set of use cases clarifying certain critical aspects of the desired system functionality. A number of specialized software tools for prototyping and user interface mock-up creation have been considered. The JustInMind software [14] has been found to best meet the set requirements. It offers a rich set of widgets that support almost all kinds of elements and interactions possible on a screen/page/form. The mockup screens developed have made possible the effective communication between the project support team and the DHARE UWG members. Instead of communicating via text, the mock-up screens have turned out to be most effective in conducting live demonstrations of critical user interface and system operation instances over the Internet for the UWG members to identify vague points that call for improvement and/or re-design. User interface and system operation instances that have been demonstrated to and commented upon by the UWG members include:

• E_I profile creation and updating

• E_I-E_I merging (i.e. resolving cases whereby two or more DHARE IDs are found to have been assigned to the same individual)

• E_O-E_O merging (i.e. cases where two E_O entities merge to comprise a single DHARE entity)

• E_O profile updating, including maintenance operations on the entity's organizational units, e.g.: creation, deletion, renaming, merging, splitting, etc.

• E_I identification and DHARE ID retrieval by a typical third party application (DSpace) utilizing the DHARE API interface

• etc.

The mock-up screen shown in Figure 4 is indicative of the type of system operation instances considered in detail the DHARE project support team and UWG members. The operation in question has to do with the identification and the grouping of all registered E_I entities that meet a set of professional affiliation criteria. For example, individuals registered to have worked at one of the organizational units of a given E_O entity during a given time period (start-end dates). The E_I entities identified are placed into a (logical) "basket" in order to undergo subsequent processing. The latter could be, for example, the starting of a new professional affiliation with another organizational unit within the same or another E_O, etc., etc. Evidently, the aforementioned operation comprises an intermediary processing stage to a number of operational instances, for example: E_O-E_O merging and splitting, organizational unit creation, deletion, renaming, merging, and splitting, etc.

Figure 4. Professional affiliations: E_I grouping and update operations

6. CONCLUSION

The DHARE directory comprises an act undertaken under the Hellenic Libraries Link (HEAL-Link) project titled "Advanced Services for Open Access Digital Libraries". The aim is to provide a service to third party applications that incorporate in their functionality the unique identification of academic and research entities in Greece. Services like those of the institutional and/or subject repositories [5], current research information systems [6], research bibliography systems [7], etc.

The usefulness and the need for a national service that assigns and maintains unique identifiers to academic and research entities at the national level are justified by the following:

- a large number of services (national and international) incorporate in their functionality the unique identification of academic ad research entities, e.g.: [15], [16], [11], [12]. A national directory (registry) of academic and research entities is to standardize entity identification across third party services that utilize incompatible identifiers. The directory can also be used as a one-stop service for its users to organize own identifiers maintained with different (inter)national services

- a number of web services can benefit from implementing the unique identification of academic and research entities (e.g. most of the institutional and subject repositories)

- DHARE is to be offered as a free service by the Hellenic Libraries Link (HEAL-Link) as compared to analogous international services that require subscription fee from member institutions (e.g. [9])

- selected academic libraries staff will have privileged access to the national directory (DHARE agents). This is to decentralize and simplify the maintenance of the directory's data content. Users in need for support (e.g. "how to" queries, deny of access, multiple profiling, etc.) will enjoy efficient and effective face-to-face service by the local library staff

- DHARE has the potential to become useful to international third party services like ORCID [9], SCOPUS [11], and InCites [13]; using the DHARE ID as a reference, the latter will access up-to-date information on academic and research entities in Greece without incorporating in their own functionality the overhead of its maintenance.

The DHARE project is currently at its design and development phase. The service is planned to start operating in late 2015.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] Directory of Hellenic Academic and Research Entities, http://dhareweb.heal-link.gr/dhare (accessed May 23, 2015)

[2] HEAL-Link, Advanced Services for Open Access Digital Libraries, http://www.seab.gr/index.php/έργα-σεαβ/129-advanced-services-seab.html (accessed May 23, 2015)

[3] Digital Convergence Operational Program, Hellenic General Secretariat for Telecommunications and Post, http://www.digitalplan.gov.gr (accessed May 23, 2015)

[4] European Regional Development Fund, http://europa.eu/legislation_summaries/employment_and_social_policy/job_creation_measures/l60015_en.htm (accessed may 23, 2015)

[5] Digital Repository, National Technical University of Athens, http://dspace.lib.ntua.gr (accessed May 25, 2015)

[6] Current Research Information System, Central Library and Information Center, University of Patras, http://cris-project.lis.upatras.gr (accessed May 25, 2015)

[7] Hellenic Researchers Index, Hellenic National Documentation Center, http://www.researchers.gr (accessed May 25, 2015)

[8] International Repositories Infrastructure wiki, https://repinf.pbworks.com/f/Author+identification+b-w.pdf (accessed May 31, 2015)

[9] Open Researcher and Contributor ID (ORCID), http://orcid.org (accessed May 31, 2015)

[10] Hellenic Academic Libraries Link (HEAL-Link), http://www.heal-link.gr (accessed May 31, 2015)

[11] Elsevier SCOPUS abstract and citation database, http://www.elsevier.com/online-tools/scopus (accessed May 31, 2015)

[12] Thomson-Reuters Web of Science, http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science.html (accessed May 31, 2015)

[13] Thomson-Reuters InCites service for research evaluation and management, http://about.incites.thomsonreuters.com (accessed May 31, 2015)

[14] JustInMind, http://justinmind.com (accessed May 31, 2015)

[15] Hellenic Academic Electronic and Interactive Books, www.kallipos.gr (accessed May 31, 2015)

[16] Academic Faculty Member Election and Promotion, https://apella.minedu.gov.gr/en (accessed May 31, 2015)

[17] ResearchGate, http://www.researchgate.net (accessed May 31, 2015