Aggregative Data Infrastructures for the Cultural Heritage

Alessia Bardi[1 2], Paolo Manghi[1], and Franco Zoppi[1]

[1] Consiglio Nazionale delle Ricerche
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
name.surname@isti.cnr.it
[2] Dipartimento di Ingegneria dell'Informazione, Università di Pisa

Abstract. Aggregative data infrastructures (ADIs) are information systems where organizations can find the tools to integrate their data sources to form uniform and richer information spaces of object metadata descriptions. Novel sustainable approaches for the realization of ADIs are based on the adoption of ADI enabling technologies which support the realization, maintenance and upgrade of ADIs and promote functionality re-use. The Cultural Heritage (CH) community is one of the most active in the realization of ADIs. Besides, the realization of ADIs for CH can be particularly complex when compared to other disciplines due to the possibly high heterogeneity of data sources involved. In this paper, we present the D-NET Software Toolkit as an ideal candidate for the realization of sustainable, extensible, scalable and dynamic ADIs for CH. To this aim we present the D-NET framework and services, and demonstrate its effectiveness in the CH scenario by describing its adoption to realize a real-case ADI for the project Heritage of the People's Europe.

Keywords: Cultural Heritage, Metadata Aggregation, Interoperability

1 Introduction

In the last decade, the multi-disciplinary character of science and the need of researchers to gain immediate access to research material often led to the realization of so-called *aggregative data infrastructures* (ADIs). These are here intended as information systems where organizations (e.g. research centers, universities, industries) can find the tools to integrate their data sources to form uniform and richer information spaces and support their communities with enhanced access services to such content. In particular, ADIs offer functionality for (*i*) the collection and processing of metadata descriptions of files (digital objects) in order to populate a uniform aggregated information space and (*ii*) the provision of the information space to humans, via web portals, and machines, via standard APIs. On the one hand, one major challenge for ADI designers and developers is to provide tools capable of dealing with several interoperability issues derived by the mismatch between the aggregated information space and the data sources; e.g. export protocols, structure and semantics of metadata, physical representation. On the other hand, another big challenge is to realize ADI capable of coping with the dynamic and complex requirements of research communities, whose needs in terms of content, functionality, and quality of service tend to vary along time, as science evolves. Indeed, software and system refinements prove to be as expensive as necessary for the ADI to grow and be up to the challenge of its community. Therefore, the adoption of the proper enabling technology plays a crucial role for the sustainability of an ADI. Such technology should minimize the cost of design and development required to realize, operate, and modify data infrastructures. The *D-NET Software Toolkit*[1] was specifically realized to facilitate designers and developers in the realization and maintenance of ADIs. D-NET implements an open source service-oriented framework where services for the collection, processing and provision of metadata and files from a set of data sources can be customized and combined to implement the internal workflows of ADIs. As proven by the several installations[3] and adoption in a number of European projects (DRIVER, DRIVER II, OpenAIRE, OpenAIREplus, EFG, EFG1914), ADIs realized with D-NET are easily customizable, extensible, scalable, and sustainable[2].

In this work we focus on the Cultural Heritage (CH) domain, which is certainly one of the most active in the realization of ADIs[3][4]. The increased availability of CH digital content raised a natural need to realize ADIs for the integration and delivery of such content to wider research, academic, and public communities[4]; examples are the ADIs supported by Europeana[4] and its satellite projects. The realization of ADIs for CH can be particulary complex when compared to other disciplines. This is due to the high degree of heterogeneity brought in by CH communities, which are typically formed by groups of subcommunities whose research focuses may diverge but require to be connected to enable better science. In this paper, we show how the D-NET Software Toolkit can be particularly apt for the realization of ADIs for CH. Besides, we propose a two-phase metadata conversion approach to tackle with the particularly complex interoperability issues which may arise in CH scenarios featuring highly heterogeneous data sources. To this aim we present the D-NET framework and services and describe their usage to instantiate a two-phase conversion ADI in the context of The Heritage of the People's Europe (HOPE) project[5]. The HOPE project provides a unified entry point for the social and labour history from the 18th to the 21st century in Europe. It federates digital object collections from several major European institutions in the field. HOPE is an exceptionally representative scenario of CH's richness, since social and labour history covers a wide range of digital objects, such as documentaries, pictures, drawings, and archival documents, in turn described by highly heterogeneous metadata representations.

Paper Outline. Section 2 presents the evolving requirements surfacing when realizing ADIs and the sustainability issues they entail for supporting organizations. Section 3 describes the architecture and functionalities of the D-NET Software Toolkit and explains how it minimizes the cost of addressing the evolving requirements of ADIs. Section 4 describes the HOPE real-case scenario and how D-NET has been successfully adopted to realize the HOPE ADI for CH.

---

[3] http://www.d-net.research-infrastructures.eu
[4] http://www.europeana.eu
[5] http://www.peoplesheritage.eu

2 Aggregative Data Infrastructures

In the last few years, an increasing number of research communities started federating their data sources into ADIs. A high-level functional architecture of ADIs is shown in Fig. 1: ADIs are inteded here as systems capable of collecting *metadata records* and *files* relative to objects from a set of heterogeneous *data sources* to construct an homogeneous information space of data conforming to a *common data model*. Over the resulting information space, the ADI provides community services to support advanced access to the aggregated data; e.g. cross-source search and browse, cross-source object interlinking, standard API exports, etc. ADIs typically focus on metadata aggregation and realize information spaces whose data can be used to cross-search over files which are kept at their original locations. In some cases however, files may be collected or uploaded in an ADI to offer services for digital preservation[5, 6]. In the following we shall describe the two main challenges to be tackled when realizing ADIs: data interoperability and curation, and coping with evolving requirements.

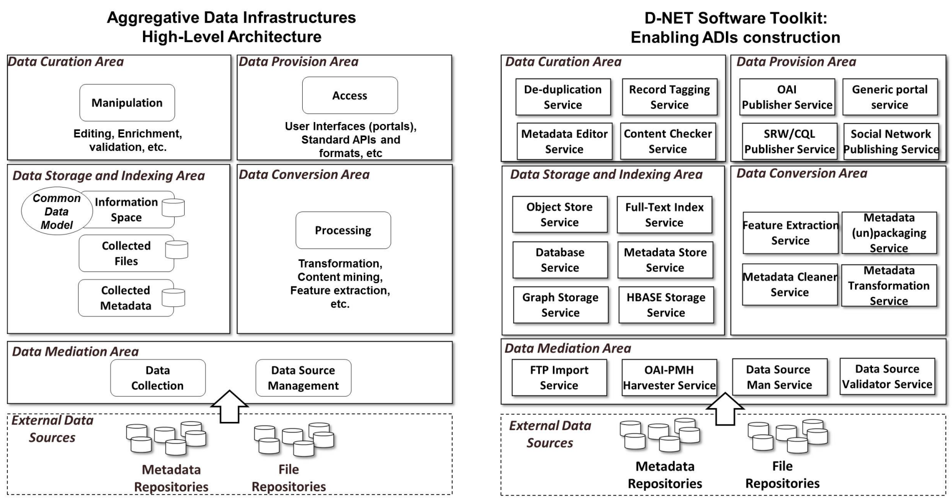

Fig. 1. Aggregative Data Infrastructures and D-NET implementation

2.1 Data interoperability and curation

ADIs collect from data sources, through standard APIs, files and relative metadata records. In the following we shall focus on XML metadata data sources, that is data sources exporting metadata records in XML format – the same or similar reasonings can be applied to all formats, e.g. RDF, JSON. Metadata records are on-the-wire representations of data conforming to the data source data model. The ADI information space contains data conforming to the given common data model whose physical representation may be based on several standard storage solutions, such as relational databases, graph stores, full-text indices, XML native stores, etc. ADIs must therefore provide tools to overcome two main interoperability barriers: the definition of logical mappings from data source data models onto ADI common data model and the definition of physical mappings from XML metadata records to ADI storage data representation. The design and implementation of ADIs must face the following technical challenges[2]:

Mediation interoperability: data sources may export data according to different standard protocols. Typically, ADIs solve this issue by natively supporting standard exchange protocols, such as OAI-PMH, FTP, HTTP, and including services capable of collecting and storing data locally.

Representation interoperability: as mentioned above collected metadata records are encoded in XML while data in the ADI information space may not necessarily be stored in the same way. Conversion software must therefore encode both logical and physical mappings from XML records onto information space objects. Typically, ADIs facilitate this task by defining a common XML schema for representing the information space data model. This leaves the logical mappings at the level of XML schemas, where XSLT mappings can be flexibly and more easily defined for each data source. Physical mappings, i.e. code to transform XML records into information space objects, is written only once.

Structural and semantic interoperability of metadata: collected metadata records are encoded in XML but according to structure (XML schemas) and semantics (e.g. vocabularies, value formats) which differ from data source to data source. These depend on the data source data model, i.e. the entities and relationships used to describe or contextualize the digital objects at hand, but also on the underlying storage platform. It is common to describe the same type of digital objects with metadata records conforming to different metadata schemes. Typically, ADIs solve this issue by including services capable of mapping input XML metadata records onto XML records conforming to the common metadata schema. Such mappings (e.g. XLST scripts), which differ based on the data source, are defined by data curators who find structural (e.g. paths to paths) and semantic (e.g. vocabulary terms to vocabulary terms) correspondences.

Granularity interoperability of metadata: by *granularity* we mean the level of data model detail represented by one XML metadata record. In some cases each record represents one entity of the model (e.g. a Dublin Core record represents and describes one publication entity), in other cases it may represent more entities possibly with relationships between them (e.g. one ESE record can represent a set of entities, while a METS record may represent a graph of interrelated entities). Since structural and semantics mappings for XML records apply at the level of the individual entities, further services are required to *(un)package* records in order to single out the entities required for the mappings.

Manipulation of information spaces: the ADI common data model is typically defined to minimize information loss w.r.t. the collected data and relative data models, but also maximize the quality and richness of the generated metadata records (e.g. entity properties should rarely have missing values). In some cases, the model includes attributes whose values may be derived by extracting information from the metadata records (e.g. mining attribute values and relationships between records to infer further values or relationships) or from the files described by such records (e.g. histograms from images, keywords from text documents). Typically, ADIs solve this issues by including services capable of processing collected data to enrich the quality of records in the information space. Moreover, collecting objects from several data sources may lead to duplication of content, whenever different sources keep information about the same entities. In such cases, de-duplication actions, i.e. merge metadata records describing the same object into one, may be necessary to disambiguate the information space. To this regard, ADIs may include de-duplication services specifically devised to exploit attributes and relationships of a record to identify and (semi-)automatically merge similar records.

2.2 Coping with evolving requirements

Organizations willing to realize ADIs must be able to sustain the initial design and development cost, plus the refinement costs made necessary by further changes required by the operative ADIs. Indeed, ADIs are often characterized by highly evolving requirements in terms of content, functionality, and Quality of Service (QoS). On the content side examples are changes to the common metadata model, new mappings required to handle interoperability with new joining data sources, etc. On the functionality side examples are changes in the data management workflows (e.g. collection, conversion, storage, indexing workflow may turn into a collection, storage, conversion, indexing workflow, to make the index more efficiently re-generated on different mapping conversions), new services to integrate missing functionality, etc. On the QoS side, management of storage and index replicas may be required to ensure robustness and availability.

Most ADI enabling software in the literature are designed to tackle very precise data aggregation scenarios and can hardly be re-used in different contexts and domains, examples are the projects Multimatch[7], KEEP[6], MICHAEL[7], DARIAH[8] and CLARIN[9]. This is due to the overall absence of general-purpose software for ADIs, which leads organizations responsible for ADIs to face the high cost of realizing their ADIs from scratch and in a very pragmatic way. This often happens by integration of open source technologies and products, such as OAI-PMH aggregators (DLXS, Repox), full-text indices (Apache Lucene and Solr), XML stores (Exists), etc. As a consequence, the result are ADIs which very efficiently address their initial requirements, but involve high refinement and maintenance costs whenever the dynamic requirements described above must be satisfied. In many cases, organizations must face the trade-off between refinement costs and end-user satisfaction.

3 The D-NET Software Toolkit

In the previous section we explained how the realization of ADIs is not trivial in terms of technical expertise, development and maintenance costs. The common approach to create from scratch such infrastructures, realizing functionalities for one specific community, hardly re-usable in other contexts, and re-implementing common functionalities from the bare metal up, instead of sharing them among communities turned out to be not affordable in the majority of cases. As a reaction to such drawbacks, research in the e-Infrastructure field concentrated on the realizion of software systems specifically designed to support the creation of

[6] Keeping Emulation Environments Portable, `"http://www.keep-project.eu"`
[7] Multilingual Inventory for Cultural Heritage in Europe, `"http://www.michael-culture.org"`

ADIs[2, 10, 11]. Typically, such software systems are based on general-purpose functional patterns for data collection, processing, storage and provision in order to allow developers to realize ADIs by re-using, customizing, and pipelining functionalities into workflows to meet the specific community needs. Moreover, the underlying loosely-coupled components architectures allow developers to dynamically modify the workflows and to integrate new functionalities. Finally, a middleware with *enabling functionality* helps ADI administrator on tasks related to the QoS, such as scalability, robustness, and load balancing.

In this section we present the D-NET Software Toolkit[8] and show how it addresses the technical challenges described in Sect. 2. D-NET is an open source, general-purpose software conceived to enable the realization and operation of ADIs (initial requirements) and to facilitate their evolution in time (refinement requirements). D-NET implements a service-oriented framework based on standards, namely Web Services with SOAP and REST APIs, where ADIs can be constructed in a LEGO-like approach, by selecting, customizing, and properly combining D-NET services. The resulting ADIs are systems which can be re-customized, extended (e.g. new services can be integrated), and scale (e.g. storage and index replicas can be maintained and deployed on remote nodes to tackle multiple concurrent accesses or very-large data size) at run-time.

D-NET offers a rich and expandable set of services targeting data collection, processing, storage, indexing, curation and provision aspects. Services can be customized and combined to meet the data workflow requirements of a target user community. D-NET services can be partly or fully replicated and distributed over different servers depending on the QoS needs of the specific community. In general, multiple instances of a service increase fault tolerance, reduce the overload of each instance, and make it possible to dynamically reorganize the environment when a server is not reachable. Figure 1, presents how several D-NET services, some of which realized in the context of the HOPE project, implement the high-level architecture and functionalities:

Data Mediation Area Services in this area are capable of managing (register and de-register) a set of available external data sources and of collecting their objects. D-NET offers services for on-demand and programmatic data collection based on the following standard protocols: OAI-PMH, FTP, FTPS (FTP over SSL/TSL), SFTP (SSH File Transfer Protocol), HTTP/HTTPS.

Data Storage and Indexing Area Services in this area manage storage and access for files and metadata records. Services offer various data storage supports, abstracting over relational databases (Postgres), file (MongoDB and file system) and graph stores (Neo4J), full-text indices (Solr), NoSQL storage (HBASE), and metadata store (abstraction on top of file storage services). Developers can configure and choose the most proper storage based on the functional requirements and the common data model of the ADI at hand.

Data Convertion Area Services in this area offer functionalities to convert XML metadata records, regardless of the structure and semantics of their schemas, and files, regardless of their storage formats. The Transformation Service can be configured to transform metadata records from one schema to another (e.g. from MARC to Dublin Core) given XSLT mappings. D-NET data managers can cre-

[8] http://www.d-net.research-infrastructures.eu

ate, update, remove and re-use such mappings and configure the service to apply one format to a given input (e.g. a metadata store) at given time intervals. In particular, in the case of one-to-one mappings between XML records, the service provides an end-user interface for the aided-creation of mappings in the style of Repox[12] and MINT[9]. The Cleaner Service can be similarly configured to harmonise values in the records based on terms-to-terms mappings. Metadata record conversion is completed by services that solve granularity issues by (un)packaging XML records (one-to-many and many-to-one conversions). Finally, the Feature Extraction Service can perform information extraction from files according to given algorithms to produce values to be added to records (e.g. extraction of histograms from images).

Data Curation Area Services in this area offer functionalities for data curation and enrichment. The Content Checker Service provides data curators with tools to find mapping mistakes and semantic inconsistencies in records of the information space. Curators can commit content (from a given data source) as visible to the public when validation is successful. The Metadata Editor Service allows data curators to add, edit and delete metadata records, as well as to establish relationships between existing records, even if coming from different sources. The De-duplication Service[13] allows data curators to disambiguate the information space by merging duplicate records. The tool identifies the pairs of records candidate for merging based on a multi-sort version of the sorted neighbourhood algorithm and a record similarity function that is customizable by data curators. Finally, D-NET was extended in HOPE to include the Record Tagging Service, which allows data curators to mark a group of objects in the information space according to terms of a given vocabulary.

Data Provision Area Services in this area allow third-party applications to access objects via standard APIs. D-NET currently supports the following provision protocols: OAI-PMH (enabling harvester to access metadata records), SRW, REST and WSDL/SOAP (enabling third party applications, such as portals, to perform queries on D-NET indices). Moreover, User Interface Services can be used to automatically generate templates of portals based on the common data model used across the storage services of the ADIs. Finally, in the context of HOPE we included new D-NET services for the automatic export (and removal) of videos and pictures towards social networks whenever such objects are tagged with special labels (e.g. tags by the record tagging service, values from a vocabulary). The Social Network Publishing Services are designed to be extendible to include further publishing actions and new web destinations.

4 D-NET in the Cultural Heritage: the HOPE ADI

HOPE (Heritage of the People's Europe, FP7 EU eContentplus, grant agreement: 250549)[10] is a "Best Practice Network" for archives, libraries, museums and institutions operating in the fields of social and union history. The goal of the project is providing a unified access to materials about the European social and labour history from the 18th to 21st centuries, proposing guidelines

[9] MINT at National Technical University of Athens, Metadata Interoperability Services. "`http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki`"

[10] http://www.peoplesheritage.eu

and tools for the management, aggregation, harmonisation, curation and provision of digital CH content. Institutions joining the HOPE network benefit of an advanced, distributed ADI instantiated and maintained by ISTI-CNR (Pisa, Italy). The ADI enables them to enhance the quality and the visibility of the digital cultural objects they preserve. Moreover, the project also delivers a Shared Object Repository, external to the ADI and realized and managed by IISG in Amsterdam, Netherlands. The repository deals with the management (storage, access, and conversion) of digital files for HOPE partners who cannot afford the cost of a local object file store. It allows institutions to deposit their files and it automatically applies conversion algorithms to create files in standard formats and with sizes suitable for web dissemination.

The HOPE ADI is implemented using the D-NET Software Toolkit, by extending it to include new services such as the Record Tagging Service and the Social Network Publishing Services (as presented in the previous section) and to adopt a two-phase approach to metadata record conversion. In the following sections we shall introduce the requirements of the HOPE infrastructure, as exemplary of the CH domain, and describe how the D-NET software is today used to implement them in an efficient and sustainable way.

4.1 The HOPE Aggregative Data Infrastructure

The project HOPE includes an initial set of content provider institutions whose common need is the realization of an ADI. The community is willing to make objects files from all data sources accessible from an aggregated information space whose metadata records obey to the same HOPE common data model. The aim is to group and interlink such objects in order to establish opportunities for a new cross-country, cross-institution social history background. The ADI should be able to handle a varying number of content providers, which may be in turn deliver several data sources, each dedicated to storage of metadata records and files relative to different object typologies; e.g. an institution may offer two data sources, relative to an archive and a library. Indeed, as it often happens in the CH domain, content providers may deliver data sources whose objects belong to diverse sub-communities (in HOPE referred to as *profiles*), which in HOPE are: *library*, *archive*, *visual*, *audio video*. Although a profile marks a data source as including material of the same "semantic domain", distinct data sources may store objects of different formats (e.g. images, videos, audio, text material) and different descriptive data models and relative metadata formats. For example, librarians and archivists typically model their digital objects according to different data models and schemata (e.g. Dublin Core for libraries, and EAD for archivists), but each of them may have a variety of ways to describe their objects (e.g. libraries may also use MARC). Furthermore, data sources may export their content via several standard protocols, such as OAI-PMH, FTP, etc.

The information space is populated by collecting and converting metadata records from HOPE content providers, and curated by HOPE data curators, who can edit/correct metadata records and tag objects in order to : (*i*) classify them, based on a vocabulary of historical themes (defined as part of the HOPE data model), or (*ii*) establish which social networks they should be sent to, based on a list of social networks. Finally, the information space is searchable and browsable by end-users from the project web portal (IALHI[11] portal) and made available to Europeana and other interested service consumers via OAI-PMH APIs.

Based on the four HOPE domain profiles, the HOPE consortium defined a common metadata model and its corresponding XML schema. In order to capture the commonalities of diverse object domains and formats, the model has been defined by studying the characteristics of the four profiles from the perspective of well-established standard format in the respective field: MARCXML for libraries, EAD for archives, EN 15907 for audio video, and LIDO for visual.

As depicted in Fig. 2, seven class of entities resulted from this process. Descriptive units represent digital objects and include descriptive information about the real world object (e.g. date of creation, type of material, title). According to the identified profiles, the descriptive unit class has four subclasses containing properties that are peculiar to one specific domain. Cross-domain properties are instead defined in the descriptive unit super class. Descriptive units are related with each other via containment and sequential relationships so that it is possible to represents hierarchies of objects (for example a book with miniatures, where there is a description of the whole book - the container - and a description of each miniatures in it). A digital resource contains technical information about a digital representation of the object (e.g. the picture of one side of a coin, the digitised page of a book) and it is linked to the corresponding descriptive unit. Digital resources related to the same descriptive units can express sequential relationships, thus establishing a "reading path". Agents, places, concepts, events, and themes contextualize the object and are linked to descriptive units via relationships whose names describe the semantics of the association.

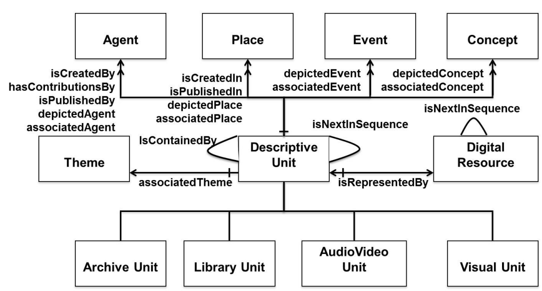

Fig. 2. HOPE common metadata model: main entities and relationships

4.2 D-NET and HOPE ADI: two-phase approach

As pointed out by Haslhofer and Klas in [14], the use of cross-walks (or mappings) solves structural and semantic heterogeneities of metadata records, enabling the realization of homogeneous information spaces where curators and automatic services can operate on. The typical approach is that of defining a common metadata format and establishing a mappings from each input format to the common one. In the case of HOPE, this process was complicated by the

[11] International Association of Labour History Institutions, "http://www.ialhi.org"

high degree of heterogeneity. As described above, since the objects and metadata records collected from the content providers may belong to sub-communities of the overall ADI, the HOPE common metadata model tends to abstract over all of such communities and therefore the mapping from data source data models into the common data model is not straightforward. For those reasons, instead of adopting a "classic" cross-walk from one input metadata format to one target metadata format, HOPE ADIs adopts a "two-phase approach". The first phase solves intra-profile structural and semantic heterogeneities, while the second phase solves inter-profile heterogeneities. The first phase is realized by mapping the metadata records of all data sources of the same profile onto metadata records conforming to a given standard data model for such profile; i.e. MARCXML (library), EAD (archive), EN 15907 (audio video), and LIDO (visual). The second phase is accomplished by providing stable mappings from such standard metadata records onto records of the HOPE data model. The approach brings two main benefits: it is easier for data source managers to map their formats into a standard format in their community (in some cases they are adopting the very same standards); and the ADI can export data source content through standard formats without further data processing. On the other side, the adoption of standards can be a drawback for data richness in cases where the input format is richer than the adopted standard. For example, multilingual descriptions may be lost when mapping onto MARCXML.

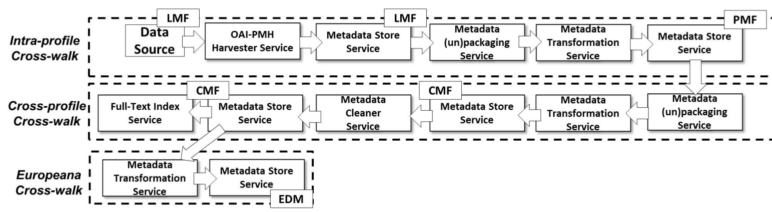

Fig. 3. The HOPE aggregation workflow

The whole aggregation workflow (Fig. 3) starts from data collection and processing and ends up to enrichment and provision. D-NET services from different functional areas are configured and combined by data managers to realize the flow, which is automatically orchestrated by the system.

Data mediation services have been configured to handle the (de-)registration and management of a variable set of data sources belonging to different content providers (organizations), each data source associated to one of the four HOPE profiles. Services can collect UTF-8 encoded XML metadata records via OAI-PMH, FTP, SFTP, HTTP and HTTPS. There are no requirements on the structure or content of records apart from one stable identifier for each record.

Data Conversion services have been combined and customized to realize the "two-phase transformation" and to deliver records to Europeana. One Tranformator Service instance is created for each data source to transform input records based on mappings defined with the help of the content provider. Once records are transformed into the profile standard format (PMF) they are further processed by another Transformator Service instance configured with the cross-walk from the PMF to the HOPE common metadata format (CMF). At this point, the flow goes to the Cleaner Service which applies the semantic transformation of values (provider vocabulary terms to common vocabulary terms). Cleaned records are delivered to the index (for use of the portal) and transformed into the Europeana Data Model[12] (EDM) to be OAI-PMH harvested by Europeana.

The HOPE information space is handled in the form of XML files, stored in Metadata Store Services and accessible via Full-text Index Services. Data curation and enrichment services, i.e. the Content Checker service and the Record Tagging Service, have also been deployed. These allow data curators to: (*i*) search and browse for records in the information space in order to check the correct implementation of the cross-walks, identify records with insufficient information, check the effectiveness of the cleansing phase, and identify records that need to be updated; and (*ii*) create new virtual, cross-data source collections by tagging records with historical themes or social network publishing tags, e.g. objects tagged with "YouTube" are automatically exported to that social site.

Data provision services export the information space via standard SRW/SRU APIs (REST and Web Services). EDM records produced in the last transformation step are published via OAI-PMH. Social Network Publishing Services have also been deployed to react based on the aforementioned tagging actions.

Today the HOPE ADI has collected 190.000 metadata records from 12 content providers, each record describing one file, converted into the HOPE common data model, and then delivered to Europeana as XML records in EDM format. At the end of the project, a total of about 900.000 metadata records will be extracted, describing around 3.000.000 files in the CH domain. HOPE digital objects will be available from the IAHLI portal and Europeana.

5 Conclusion

We highlighted the need for aggregative data infrastructures (ADIs) in the Cultural Heritage (CH) domain and described the important role of enabling software for ADIs. We argued that the realization of ADIs with a from scratch approach is not affordable in terms of re-usability, scalability and dinamicity required to satisfy the evolving functional and architectural requirements of modern user communities. On this regard, we presented the D-NET Software Toolkit, which is specifically designed to support developers of ADIs to address the above issues with a generic and effective approach, providing ready-to-use services that can be configured, extended and composed in workflows to meet the specific community's needs. We demonstrated the effectiveness of D-NET in the CH domain by describing how it has been adopted in the context of the HOPE project for the realization of an ADI implementing a two-phase approach to metadata record conversion.

References

1. Manghi, P., Mikulicic, M., Candela, L., Artini, M., Bardi, A.: General-Purpose Digital Library Content Laboratory Systems. In Lalmas, M., Jose, J., Rauber, A., Sebastiani, F., Frommholz, I., eds.: Research and Advanced Technology for Digital Libraries. Volume 6273 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, Berlin, Heidelberg (2010) 14–21–21

[12] http://pro.europeana.eu/edm-documentation

2. Manghi, P., Mikulicic, M., Candela, L., Castelli, D., Pagano, P.: Realizing and maintaining aggregative digital library systems: D-net software toolkit and oaister system. D-Lib Magazine **16**(3/4) (2010)
3. Blanke, T.: From tools and services to e-infrastructure for the arts and humanities. In Lin, S.C., Yen, E., eds.: Production Grids in Asia. Springer US (2010) 117–127
4. Loebbecke, C., Thaller, M.: Digitization as an it response to the preservation of europe's cultural heritage. In Carugati, A., Rossignoli, C., eds.: Emerging Themes in Information Systems and Organization Studies. Physica-Verlag HD (2011) 359–372
5. Cramer, T., Kott, K.: Designing and implementing second generation digital preservation services: A scalable model for the stanford digital repository. D-Lib Magazine **16**(9/10) (2010)
6. Li, Y., Banach, M.: Institutional repositories and digital preservation: Assessing current practices at research libraries. D-Lib Magazine **17**(5/6) (2011)
7. Amato, G., Debole, F., Peters, C., Savino, P.: Multimatch: Multilingual/multimedia access to cultural heritage. In Agosti, M., Esposito, F., Thanos, C., eds.: IRCDL, DELOS: an Association for Digital Libraries / Department of Information Engineering of the University of Padua (2009) 162–165
8. Blanke, T., Hedges, M.: Providing linked-up access to cultural heritage data. Proceedings of the ECDL 2008 Workshop on Information Access to Cultural Heritage. Aarhus, Denmark (2008)
9. Váradi, T., Krauwer, S., Wittenburg, P., Wynne, M., Koskenniemi, K.: Clarin: Common language resources and technology infrastructure. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., Tapias, D., eds.: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco, European Language Resources Association (ELRA) (may 2008) http://www.lrec-conf.org/proceedings/lrec2008/.
10. Simeoni, F., Candela, L., Lievens, D., Pagano, P., Simi, M.: Functional adaptivity for digital library services in e-infrastructures: the gcube approach. In Agosti M., Borbinha J., K.S.P.C.T.G., ed.: Research and Advanced Technology for Digital Libraries. 13th European Conference, Corfu, Greece, Springer Verlag (2009) 51 – 62 In: ECDL 2009 - Research and Advanced Technology for Digital Libraries. 13th European Conference (Corfu, Greece, 27 September - October 2 2009). Proceedings, pp. 51 - 62. Agosti M., Borbinha J., Kapidakis S., Papatheodorou C., Tsakonas G (eds.). (Lecture Notes in Computer Science, vol. 5714). Springer Verlag, 2009.
11. Calanducci, A., Sevilla, J., Barbera, R., Andronico, G., Saso, M., De Filippo, A., Iannizzotto, S., Vicinanza, D., De Mattia, F.: Cultural heritage digital libraries on data grids. In Agosti, M., Borbinha, J., Kapidakis, S., Papatheodorou, C., Tsakonas, G., eds.: Research and Advanced Technology for Digital Libraries. Volume 5714 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2009) 469–472
12. Reis, D., Freire, N., Manguinhas, H., Pedrosa, G.: Repox – a framework for metadata interchange. In Agosti, M., Borbinha, J., Kapidakis, S., Papatheodorou, C., Tsakonas, G., eds.: Research and Advanced Technology for Digital Libraries. Volume 5714 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2009) 479–480
13. Manghi, P., Mikulicic, M.: Pace: A general-purpose tool for authority control. In García-Barriocanal, E., Cebeci, Z., Okur, M., Öztürk, A., eds.: Metadata and Semantic Research. Volume 240 of Communications in Computer and Information Science. Springer Berlin Heidelberg (2011) 80–92
14. Haslhofer, B., Klas, W.: A survey of techniques for achieving metadata interoperability. ACM Comput. Surv. **42**(2) (March 2010) 7:1–7:37