On Bridging Data Centers and Publishers: the Data-Literature Interlinking Service

Adrian Burton[1], Hylke Koers[2], Paolo Manghi[3], Sandro La Bruzzo[3], Amir Aryani[1], Michael Diepenbroek[4], Uwe Schindler[4]

[1] Australian National Data Service,
Melbourne, Australia
{adrian.burton, amir.aryani}@ands.org.au
[2] Elsevier,
Amsterdam, The Netherlands
{h.koers}@elsevier.com
[3] Institute of Information Science and Technology - CNR,
Pisa, Italy
{paolo.manghi, sandro.labruzzo}@isti.cnr.it
[4] PANGAEA,
Bremen, Germany
{mdiepenbroek, uschindler}@pangaea.de

Abstract. Although research data publishing is today widely regarded as crucial for reproducibility and proper assessment of scientific results, several challenges still need to be solved to fully realize its potential. Developing links between the published literature and datasets is one of them. Current solutions are mostly based on bilateral, ad-hoc agreements between publishers and data centers, operating in silos whose content cannot be readily combined to deliver a network connecting research data and literature. The RDA Publishing Data Services Working Group (PDS-WG) aims to address this issue by bringing together different stakeholders to agree on common standards, combine links from disparate sources, and create a universal, open service for collecting and sharing such links: the Data-Literature Interlinking Service. This paper presents the synergic effort of the PDS-WG and the OpenAIRE infrastructure to realize and operate such a service. The Service populates and provides access to a graph of dataset-literature links collected from a variety of major data centers, publishers, and research organizations. At the time of writing, the Service has close to one million links with further contributions expected. Based on feedback from content providers and consumers, PDS-WG will continue to refine the Service data model and exchange format to make it a universal, cross-platform, cross-discipline solution for collecting and sharing dataset-literature links.

1 Introduction

Driven by innovations in digital technology and off-the-shelf availability of cheap storage solutions, research data is becoming ever more prominent in the way that research is performed and in the way research findings are communicated. Research

data holds a big promise, and improving the storing, sharing, and usage of data is seen by many as a powerful way to accelerate the pace of science, even fuel economic growth. As Neelie Kroes, then Vice-President of the European Commission responsible for the Digital Agenda put it: "Knowledge is the engine of our economy. And data is its fuel."

Challenges to realize the full potential of research data exist at different levels - from cultural aspects, such as proper rewards and incentives, to policy and funding, and to technology. The challenges are interconnected and impact a diversity of stakeholders - including researchers, research organizations, funding bodies, data centers, and publishers. It is essential that these stakeholders work together to address common issues and push the envelope. ICSU World Data Systems (ICSU-WDS) and the Research Data Alliance (RDA) provide useful forums for such collaborations, such as the Publishing Data Interest Group (IG). This IG addresses a range of issues in data publication from a holistic and cross-stakeholder perspective, acting as the umbrella of Working Groups (WGs) that deal with data bibliometrics, data publication workflows, cost recovery, and services. Among these WGs, the Publishing Data Services WG (PDS-WG) brings together different parties in the research data landscape (e.g. data centers and publishers) with the objective of creating "an open, freely accessible, web based service that enables its users to identify datasets that are associated with a given article, and vice versa" [1]. The vision is that of moving away from the large set of bilateral arrangements that characterizes the research eco-system today, towards establishing common standards and tools that sit in the middle and interact with all parties. Such a transition would facilitate interoperability between platforms and systems operated by the different parties, reduce systemic inefficiencies in the ecosystem, and ultimately enable new tools and functionalities to the benefit of researchers.

This paper presents the work carried out by PDS-WG in the realization of a Data-Literature Interlinking Service (referred to as "the Service" in the following) capable of supporting such a shift. In this process, the WG has joined forces with the OpenAIRE project[1] and infrastructure [10] in order to design, develop and deploy an operative and sustainable prototype of the Service. The Service has been conceived in such a way that its common data model and exchange format can be refined over time to become community-driven standards, balancing between the information that can be shared across data providers and the information that is needed by consumers of the Service.

2 The need for a Data-Literature Interlinking Service

The most immediate benefit in establishing links between articles and data is to increase visibility and discoverability, thus bringing data (and articles) more to the forefront and stimulating re-use. In addition, by providing links to the scholarly literature, data can be put in the right context that is often necessary to reproduce findings

[1] OpenAIRE, http://www.openaire.eu

or re-use data properly (see also [5]). Researchers across disciplines strongly support the notion that there is value is creating links between data and the literature, as testified by results from the PARSE.Insight study[2], which was carried out with the help of EU funding in 2008—2010 : 85% respond *"yes"* to the question *"Do you think it is useful to link underlying research data with formal literature"* [5]. However, what is also clear is that in order to be fruitful, such linking needs to be done properly, by means of infrastructural solutions, delivering agreed-on policies, formats, and tools [3]. For example, a recent study in the astronomical literature showed that more than 50% of links from articles to data using a hard-coded HTTP web address were broken after 15 years [6]. Many parties, in fact, are taking efforts to link up articles and data in a robust and future-proof way: A number of data repositories keep track of articles that cite, or refer to, their data; several publishers have some form of data-linking program to connect the articles they publish with relevant data hosted externally (see e.g. [7]); providers of bibliographic information are increasingly looking at data alongside the traditional article output; and organizations such as CrossRef, DataCite and OpenAIRE are developing systems to track or infer relationships between data and the literature (see also [8] for some examples of how data and literature publications are currently interlinked).

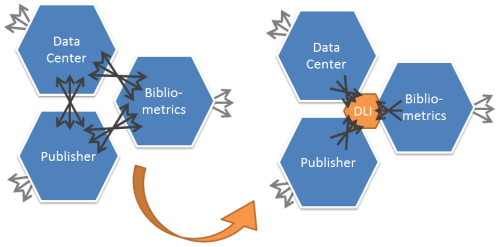


Figure 1 - Moving towards common standards and one-for-all services.

However, these initiatives typically live in isolation, and there is no common framework for inter-linking datasets and published articles. As a consequence, although different parties have a "piece of the puzzle", those pieces cannot be readily combined to exploit at best the potential of a rich and comprehensive network of published literature and data sets. The work of PDS-WG is seeking to tackle the comprehensiveness and interoperability challenges underlying this scenario by realizing an open and one-for-all Data-Literature Interlinking Service (**Figure 1**). The Service will serve as a flexible sandbox where major scholarly communication stakeholders interested in sharing or consuming dataset-literature links will be able to do so while reporting their requirements, preferences, recommendations, obstacles to the PDS-WG. Such an incremental approach will enable the refinement of exchange formats, data model, and aggregation workflows implemented by the Service and, in the long run, to agree on common practices for sharing dataset-publication links. The operation of the Service will bring the following benefits (adapted from [1]):

[2] *PARSE.Insight project*, http://www.parse-insight.eu/

1. *For data repositories and journal publishers*: it will make the process of linking data sets and research literature more scalable and with less overhead, ensuring more visibility for data sources (and their "customers") as well as publication platforms.
2. *For research institutes, bibliographic service providers, and funding bodies*: it will enable advanced bibliographic services and productivity assessment tools that track datasets and journal publications within a common framework;
3. *For researchers*: it will make the processes of sharing and accessing relevant articles and data easier, more efficient, and more accurate, thereby increasing scientific reward and enhancing its practices.

2.1 Modus operandi

Four key principles underpin the thinking and the work carried out in the PDS-WG. First, the challenge of developing an open, universal interlinking system is as much of a "soft" (social) problem as it is a "hard" (technical) problem. The WG has therefore invested a considerable amount of time and effort in building a broad base of support through communication and outreach activities. Today all of the groups that were identified as key stakeholders - including data centers, publishers, providers of bibliographical information, funding bodies, etc. - are supporting the initiative, be it through WG membership, contributing a corpus of article/data links, participating in the technical work, or a combination thereof. The initiative is open and inclusive[3] and additional participation by other groups or individuals will be welcomed.

Second, the WG is prioritizing its efforts towards building, a working prototype of the Service that can be used to demonstrate value to the intended users and stakeholder groups. This work is carried out in synergy with the OpenAIRE project and infrastructure, PANGAEA, and ANDS. As with any demonstrator system, coverage and functional scope are initially limited but the ambition is to develop a service that will be of direct value in real-world situations. The admittedly important set of questions around longer-term sustainability and governance of the Service is deferred to a later stage of the WG's lifetime. Specifically, a pragmatic, ground-up approach was followed: aggregate as many corpora of literature-data links as possible, harmonizing them into a common data model, and making them available online through an openly accessible Service.. That means that in the initial stage of operation the WG admits a considerable effort to ingest heterogeneous information from contributors. In the long run, the expectation is that the Service will help at establishing exchange standards that will reduce conversion costs and lead to a more scalable approach. To this aim the Service will enable a "test & learn" approach, by facilitating the extension of the common data model and schema over time.

[3] A set of "guiding principles" that includes statements on the open character of the project can be accessed through the WG's RDA website: https://www.rd-alliance.org/groups/rdawds-publishing-data-services-wg.html

Third, the WG takes a generic, one-size-fits-all (as opposed to e.g. domain-specific) approach as much as possible to avoid fragmentation and preserve the value that lies in developing a comprehensive solution for all articles and all datasets. This approach necessarily means that the Service common data model is relatively discipline-agnostic, leaving domain-specific metadata a responsibility of the data repositories.

Finally, the WG places significant emphasis on provenance, reliability, quality of data-literature links and the associated metadata, considered of great importance for most key use cases (e.g. linking from online publishing or data platforms, bibliometrical analyses). This principle is reflected in the Service operation, which ensures that: (i) links are contributed by trusted sources, rather than inferred by the system, and (ii) the origin and completeness of links and metadata is tracked at a high level of detail and granularity.

2.2 Related Work

The ambition to develop a Data-Literature Interlinking Service is not unique, and there are a number of related initiatives. In particular, CrossRef and DataCite have announced they will be working on increasing interoperability between their systems to more easily expose article/data links in cases where both can be identified through DOI's. Other initiatives – though often broader in scope than "just" linking literature and data, for example including funder or researcher ID's – include the RMap project [11], the National Data Service[4], bioCADDIE[5], the Open Science Framework[6], and THOR[7]. In addition, there are several RDA WGs and IGs for which data-literature linking is also an important topic, most notably the Data Description Registry Interoperability (DDRI)[8] WG, which has developed RD-Switchboard.org[9]. Apart from its own development agenda, the PDS-WG aims to provide a forum for such initiatives to share and discuss their ideas, so as to avoid duplication, learn from each other, and cooperate.

3 The Data-Literature Interlinking Service

The Data Literature Interlinking (DLI) service ("the Service") aims to populate and provide access to the *DLI information space*, a graph of relationships between datasets and the literature, and between datasets and datasets. Objects and relationships are provided by data sources managed by publishers (e.g. Elsevier, Thomson Reuters), data centers (e.g. PANGAEA, CCDC), or other organizations providing services to manage links between datasets and publications (e.g. DataCite, OpenAIRE). The

[4] National Data Service, http://www.nationaldataservice.org/
[5] BioCADDIE, https://biocaddie.org/
[6] https://osf.io/
[7] THOR EC project, http://cordis.europa.eu/project/rcn/194927_en.html
[8] *DDRI*, https://www.rd-alliance.org/group/data-description-registry-interoperability.html
[9] See http://www.rd-switchboard.org/

Service aggregates content and implements programmatic access (APIs) to the resulting information space. Such APIs offer full-text search by field or free keywords and bulk access to the collection (e.g. OAI-PMH protocol). They enable the construction of services on top of the Service (e.g. end-user search and statistics portal) and serve content to third-party community services (e.g. RD-Switchboard).

The Service is also intended as a flexible playground where data curator users can monitor the aggregation outputs, collect feedback from data providers and service consumers, and refine ingestion workflows and common data model accordingly. The expectation is that such incremental and agile methodology will converge to an ideal data model and exchange metadata format for description and exchange of links between datasets and publications. The following sections present the functional requirements of the Service and the initial DLI information space data model.

3.1 Functional requirements

Users of the Service. The Service will support four categories of users. *Data source managers*, serving content to the Service and willing to gain visibility and serve their user communities; *Portal end users* (e.g. researchers, funders, publishers, data centres), searching for datasets or publications via their relationships or for statistics regarding the provenance of objects and relationships; *Service data curators*, needing tools to monitor and orchestrate their data aggregation activities in order to guarantee an expected QoS; and *Third-party service developers*, willing to (bulk) collect the DLI information space to process and offer it to their users.

Aggregating content from data sources. Data sources are intended as providers interested in feeding object-to-object relationships to the Service. Data sources deliver to the service so-called *metadata packages* (records) that encode the description of how a *source object* is interlinked via relationships to a set of *target objects*; objects are uniquely identified by a PID together with its namespace (e.g. DOI, PMCID, URL). Data sources can provide metadata packages according to three modalities: *pull*, i.e. the Service bulk-collects relationships via data source APIs; *push*, i.e. the data source sends relationships into the Service; or *resolution*, i.e. the Service collects content about one object and its relationships sending a PID to a data source resolver service. Data source resolvers (e.g. DataCite, CrossRef, PDB) are used to complete object metadata when this was not fully included in its original metadata package.

In the future, data sources should deliver metadata packages that conform to an exchange format and data model recommended by the DLI information space, the exchange format being entitled to become a standard for sharing dataset-literature links. In the initial stage of operation, the Service cannot expect data source to conform to such format. It must therefore provide mechanisms to map metadata packages, whatever native data model and format they conform to, onto objects and relationships conforming to the DLI information space data model.

De-duplication. Different data sources may provide duplicate information about the same objects and relationships: objects with the same PID-namespace from different sources or objects with different PID-namespaces (e.g. DOIs and PMCIDs) but corresponding to the same dataset or publication. The service will deliver de-duplication tools, capable of identifying groups of duplicates and merging them into one "representative" object. Representative objects will keep the PID-namespaces of the objects they merge and maintain a reference to their original data sources.

Publishing the information space graph. The Service provides a web portal for end users to (full-text) search and browse relationships between datasets and publications and to visualize statistics on the distribution of such relationships (e.g. per data source, per type, etc.). Moreover, it supports OAI-PMH APIs to export the DLI information space towards interested third-party services. Looking ahead, the PDS-WG is working to connect the Service with a data-linking provision platform developed by PANGAEA, and with an interactive network visualization tool developing in the context of RD-Switchboard (this will be discussed in more detail in section 4.3).

3.2 Data model

The conceptual data model of the DLI information space is depicted in **Figure 2**. The model (as well as the corresponding exchange format defined in the following section) is intended as an initial starting point, but is bound to be refined, as new requirements from service stakeholders and consumers will surface. Objects can be of two types, *publications* (intended as scientific literature) or *datasets*.[10] Relationships between them are directed and bidirectional; e.g. if an object A has a relationship "isCitedBy" to an object B then also the inverse relationship "cites" will be found in the information space. Relationships bear semantics, expressed by a label that belongs to a given ontology (e.g. DataCite vocabulary), and may contain a *description* in order to encode and represent dataset annotations.

Items (i.e. objects and relationships) are into the system because either (*i*) they have been pulled from external providers, (*ii*) pushed by third-party services, or (*iii*) obtained by resolving a PID using a resolver service. In order to keep track of their provenance, items are equipped with *provenance information* that consists of a reference to the originating data source, the time of ingestion of the item into the system, and the modality of bringing the item into the system ("pull", "push", "resolved"). The field *completion_status* in object provenance tracks down whether the data source has contributed full object metadata description or only a PID-namespace. This way the Service can identify which objects are "incomplete" and should be subject to subsequent resolution attempts. When the same items are provided by different data sources (duplicates) and are merged together into one representative item to disambiguate the information space, then the resulting "representative" item keeps provenance information about all the items it merges.

[10] Currently, only title, authors and publication date fields are kept, but this choice may change in the future based on user or third-party service needs.

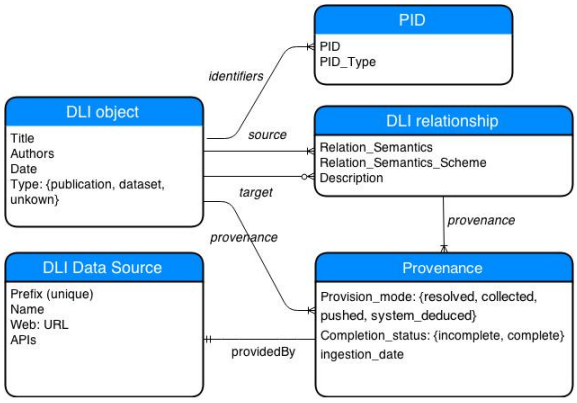


Figure 2 – Conceptual Data Model

4 The Prototype

The Service prototype is powered by the D-NET Software Toolkit ([2]). D-NET is today the platform of production systems of several aggregation infrastructures (e.g. OpenAIRE, EFG/EFG1914, HOPE, EAGLE) and repository federations (e.g. CEON Poland, MINCYT Argentina, FECYT Spain). The software is devised to enable the construction and monitoring of aggregative data infrastructures, by orchestrating a set of highly configurable D-NET data aggregation services (and/or third-party web services) into autonomic workflows. For example, data storage is possible via relational databases (Postgres), XML databases (Exist), column stores (MongoDB, HBASE), full-text indices (Apache Solr) and remote file systems (GridFS); while data processing is available via general purpose services, such as XSLT engines, Groovy Engines, Hadoop MapReduce, which are highly configurable and already embed customizable algorithms for metadata transformation, de-duplication, and inference by (text)mining collected files or metadata.

The Service prototype implements an *aggregation system* and a *provision system* as described in the following sections. The prototype meets all requirements described in the previous sections, except for the following: (*i*) data sources are only of type "pull" and "resolution"; (*ii*) de-duplication is implemented only at the level of PID-namespace (i.e. provenance information does not include original PIDs), and (*iii*) the semantics of relationships is limited to the subset of DataCite (i.e. no support for multiple vocabularies): *references, cites, isSupplementTo, isReferredBy, isCitedBy, isSupplementedBy* and otherwise mapped onto the *unknown* value.

4.1 Content aggregation system

The system handles (de)registration of data sources and aggregation of their content. Data sources register to the Service by submitting a profile describing their general properties (e.g. name, location, etc.) and technical interoperability properties (e.g. data

collection APIs, data collection modality). Each registered data source is associated with an autonomic workflow (see Figure 3) that collects its metadata packages and processes them to populate the DLI information space.

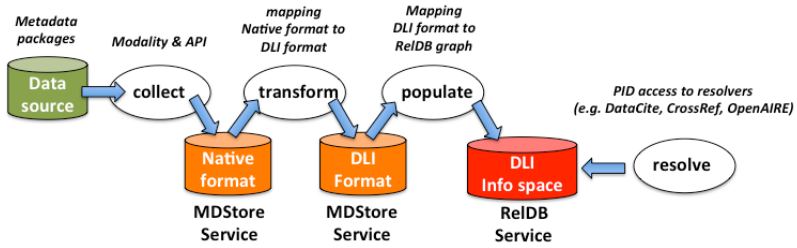


Figure 3 - Data source aggregation workflow.

To achieve this, the workflow makes use of D-NET's MetadataStore Service, Transformation Service, and RelationalDB Service. Initially, metadata packages are cached in their native format (e.g. XML, CSV, TXT), then transformed, given a set of mapping scripts, from such format onto an internal XML format called "DLI" (Table 1).

Table 1 - DLI record structure

```
DLI_ID: % obtained as <PID_type>::<PID>
PID
PIDType: % from a vocabulary doi, PMCID, ncbin, pdb, etc.
authors
title
date
type: {publication, datasets, unknown}
provenance*
    providedBy_datasource
    provision_mode: {resolved, collected, pushed, system_deduced}
    ingestion_date
    completion_status: {incomplete, complete, failed_to_resolve}
        % incomplete => type, authors, title, and date fields
        % are empty
relationship*
    target_object_type: {publication, dataset, unknown}
    target_object_title  % to be used as anchor label
    target_object_PID:
    target_object_PIDType % doi, PMCID, others
    target_object_DLI_ID
    provenance*
        providedBy_datasource
        provision_mode: {resolved, collected, pushed,system_deduced}
        completion_status: {incomplete, complete, failed_to_resolve}
        ingestion_date
        relationship_completion_status: {incomplete, complete}
            %  incomplete =>  type and title fields are empty
        semantics
            % from DataCite relationships vocabulary or "unknown"
```

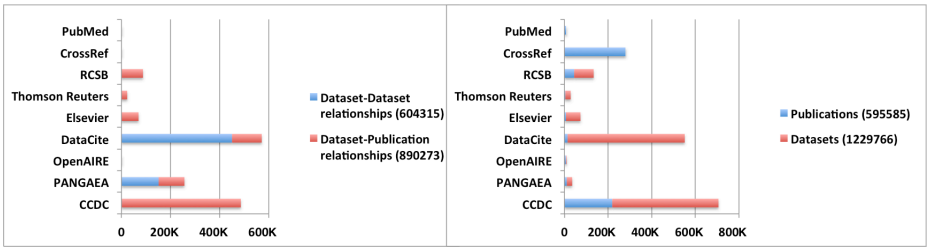
The DLI exchange format includes all information described in the data model, but also introduces some redundancy in order to become self-explanatory (e.g. enabling interpretation of target objects without necessarily accessing them). A conclusive step will transform DLI records into objects and relationships of the graph, which are encoded as records of a relational database. The graph thus built may feature objects

whose *completion_status* is "*incomplete*". Accordingly, whenever an ingestion workflow terminates, the Service fires a resolution workflow, which finds such objects, identifies the respective resolver service based on the object PID namespace, and tries to fetch the missing metadata fields. The result of such operation, be it successful or not, is tracked by the system and ends up enriching the provenance information of the given objects.

4.2 Content provision system

The content provision system consists of a workflow that is fired whenever data ingestion and resolution workflows are terminated. The workflow collects the information space graph from the RelationalDB Service, converts its objects onto DLI exchange format records (post duplicate identification and object resolution), and delivers them (in parallel) to a Full-text Index Service and an OAI-PMH Publisher Service. Users can search and browse the index from a portal available at http://dliservice.research-infrastructures.eu, while OAI-PMH APIs are available from http://dliservice.prototype.research-infrastructures.eu/oai. Currently, the prototype includes relationships and objects from the data sources reported in **Table 2**.

Table 2. Objects and relationships contributed by data source at the moment of writing. At the time of writing, the service holds 890273 dataset-literature links; further contributions from these and other sources to the PDS-WG are expected.



4.3 Forthcoming actions

The prototype will be completed to allow "push" modality for data sources, introduce de-duplication across different PID-namespaces (using D-NET de-duplication Services [4]), modify the model in order to introduce relationships of type "annotations", and support more advanced access modalities to the Information Space. On this last matter, the PDS-WG is working to connect the Service to a linking service backbone under development by PANGAEA and to leverage network visualization tools developed in the context of the RD-Switchboard platform for a front-end demonstrator tool that allows users to explore the literature-data network. Finally, we expect additional contributions from organizations represented in the PDS-WG to substantially increase the number of literature-data links.

Upgrading to PANGAEA provision system. The PANGAEA data center team is working to extrapolate the current PANGAEA linking service[11] into a generally usable linking service that will enhance the current Service content provision system. The service will offer PID-resolution APIs and be optimized for high-volume read access by science publishers and bibliometrics service providers. It will be based on Elasticsearch[12], hosted in the Amazon EC2 cloud, and will provide linking information and render metadata badges that can be embedded into article publisher's web pages to show linked data sets (see [7]). Based on this service, a new section of the portal will display linking statistics based on Elasticsearch aggregations using visualization features of Kibana[13].

Integration with RD-Swichboard. RD-Switchboard is an interoperability platform developed by ANDS in the context of DDRI-WG of RDA (Data Description Registry Interoperability), whose aim is to offer cross-platform discovery of related research datasets. The platform aggregates links between publications, datasets and research grants from national and international data services/centers (members of the DDRI-WG); then, it adopts graph-modeling techniques (e.g. exploiting co-authorship or related research projects) to identify missing links between related works. For example, RD-Switchboard has identified the datasets co-authored by Australian researchers in Dryad and CERN data repositories, and linked them to datasets in the Research Data Australia repository.

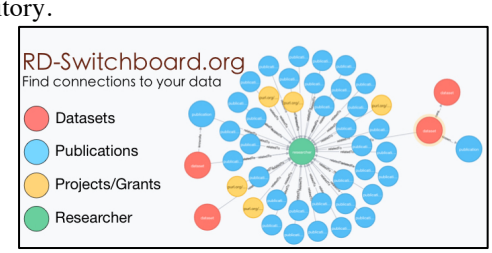


Figure 4 – Screenshot of the RD-Switchboard browser

As a result of the integration, the Service will benefit from RD-Switchboard's graph navigation and visualization tools (Force Directed Graph Drawing Algorithm [9], see Figure 4), while RD-Switchboard will profit from the rich set of literature-data links.

5 Conclusions

This paper described the work carried out in the context of the joint ICSU-WDS and RDA Working Group "Publishing Data Services" (PDS-WG) with the support of OpenAIRE. This WG has set out to create an open, universal Data-Literature Inter-

[11] *Elsevier and PANGAEA Take Next Step in Connecting Research Articles to Data,* http://www.prnewswire.com/news-releases/elsevier-and-pangaea-take-next-step-in-connecting-research-articles-to-data-99533624.html. See also [7].
[12] Elasticsearch, https://www.elastic.co/products/elasticsearch
[13] Kibana, https://www.elastic.co/products/kibana

linking Service that aggregates, harmonizes, completes, and offers access to links between the scholarly literature and research data. The technical development reflects the WG's principles of openness, inclusivity, quality, provenance, domain-agnosticism – and, finally, a pragmatic, "ground-up" approach to develop software in a test-and-learn approach that allows for continuous refinement of the system and the underlying data model. By establishing this service, the PDS-WG aims to progress the field from the current situation of many ad-hoc, bilateral agreements (between e.g. a data center and a publisher) to realize a one-for-all service architecture with common standards to the benefit of all stakeholders in the research data landscape.

Acknowledgements. The authors would like to thank the PDS-WG members and representatives from CrossRef, DataCite, The National Data Service, ORCID, The Research Data Alliance, ICSU World Data Systems, and the RMap project for many valuable discussions and constructive interactions. This work is partially funded by the EU projects RDA Europe (FP7-INFRASTRUCTURES-2013-2, grant agreement: 632756) and OpenAIRE2020 (H2020-EINFRA-2014-1, grant agreement: 643410).

References

1. *Publishing Data Services Working Group Case Statement,* https://www.rd-alliance.org/filedepot/folder/114?fid=239
2. Manghi, P., Artini, M., Atzori, C., Bardi, A., Mannocci, A., La Bruzzo, S., Candela L, Castelli D, Pagano, P. (2014). The D-NET software toolkit: A framework for the realization, maintenance, and operation of aggregative infrastructures. Program: electronic library and information systems, 48(4), 322-354.
3. Castelli, D., Manghi, P., & Thanos, C. (2013). A vision towards scientific communication infrastructures. International Journal on Digital Libraries, 13(3-4), 155-169.
4. Manghi, P., Mikulicic, M., Atzori, C. (2012). De-duplication of aggregation authority files. International Journal of Metadata, Semantics and Ontologies, 7(2), 114-130.
5. Smit, E. (2011). Abelard and Héloise: Why Data and Publications Belong Together. D-Lib Magazine, volume 17. DOI: 10.1045/january2011-smit
6. Pepe, A., Goodman, A., Muench, A., Crosas, M., Erdmann, E. (2014). How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers. PLOS One. DOI: 10.1371/journal.pone.0104798
7. Aalbersberg IJ. J., Dunham J., Koers H. (2011). Connecting Scientific Articles with Research Data: New Directions in Online Scholarly Publishing. Proceedings of the 1st ICSU World Data Systems Conference.
8. Callaghan, S., Tedds, J., Lawrence, R., Murphy, F., Roberts, T., Wilcox, W. (2014). Cross-Linking Between Journal Publications and Data Repositories: A Selection of Examples. International Journal of Digital Curation. DOI: 10.2218/ijdc.v9i1.310
9. Kobourov, Stephen G (2012). Spring embedders and force directed graph drawing algorithms. *arXiv preprint arXiv:1201.3011* (2012).
10. Manghi, P., Bolikowski, L., Manold, N., Schirrwagen, J., & Smith, T. (2012). Openaireplus: the european scholarly communication data infrastructure. D-Lib Magazine, 18(9), 1.
11. *The RMAP project, white paper, http://rmap-project.info/rmap/wp-content/uploads/RMap_Project_Overview_Revised_Final.pdf*