

Grant agreement no. 312788



**ORCID AND DATACITE  
INTEROPERABILITY NETWORK**

<http://odin-project.eu>

## **D6.1 International validation of D3.1, D3.2, D4.1 and D5.1**

**WP6 Internationalisation**

**V1.4**

Abstract: This report (D6.1) examines and validates the ODIN approach from an international perspective. In addition, it presents the feedback of the Australian and US partners on the ODIN deliverables D3.1, D3.2, D4.1 and D5.1.



**Lead beneficiary:** Australian National Data Service

**Date:** 14/10/2013

**Nature:** Report

**Dissemination level:** PU (Public)



	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final 2/13

## Document Information

<b>Grant Agreement no.</b>	312788	<b>Acronym</b>	ODIN
<b>Full title</b>	ORCID and DataCite Interoperability Network		
<b>Project URL</b>	<a href="http://odin-project.eu">http://odin-project.eu</a>		
<b>Project Coordinator</b>	John Kaye (BL) Address: The British Library 96 Euston Road, London NW1 2DB, United Kingdom Phone: +44 20 7412 7450 Email: <a href="mailto:john.kaye@bl.uk">john.kaye@bl.uk</a>		

<b>Deliverable</b>	<b>Number</b>	6.1	<b>Title</b>	International validation of D3.1, D3.2, D4.1 and D5.1
<b>Work package</b>	<b>Number</b>	6	<b>Title</b>	Internationalisation

<b>Document identifier</b>	ODIN-WP6-International_validation-0001-1_0			
<b>Delivery date</b>	<b>Contractual</b>	Month 12	<b>Actual</b>	13 <sup>th</sup> October 2013
<b>Status</b>	Version 1_4		Final <input checked="" type="checkbox"/> Draft <input type="checkbox"/>	
<b>Nature</b>	Report <input checked="" type="checkbox"/> Prototype <input type="checkbox"/> Demonstrator <input type="checkbox"/> Other <input type="checkbox"/>			
<b>Dissemination Level</b>	<input checked="" type="checkbox"/> Public <input type="checkbox"/> Restricted to other programme participants (including the Commission Services) <input type="checkbox"/> Restricted to a specified group (including the Commission Services) <input type="checkbox"/> Confidential, only for consortium members (including the Commission Services)			

<b>Authors (Partner)</b>	Amir Aryani (ADA), Todd Vision (Dryad), Simeon Warner (Arxiv)			
<b>Responsible Author</b>	Amir Aryani		<b>Email</b>	amir.aryani@ands.org.au
	<b>Partner</b>	Australian National Data Service	<b>Phone</b>	+61 2 6125 0586

	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final 3/13

<b>Document Status Sheet</b>			
Issue	Date	Comment	Author
0.1	04/10/13	Initial Draft	Amir Aryani
0.2	08/10/13	Review (minor changes)	Adrian Burton
0.21	10/10/13	Review	Laure Haak
0.22	10/10/13	Review	John Kaye
1.0	12/10/13	Merged and fixed prior comments	Amir Aryani
1.1	12/10/13	Fixed headers and footers	Amir Aryani
1.2	13/10/13	Improved the discussion about ORCID	Amir Aryani
1.3	13/10/13	Minor formatting changes	John Kaye
1.4	14/10/13	Formatting & comments by CERN	Amir Aryani

### Document Change Record

Issue	Item	Reason for Change

	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>	<b>Dissemination level: PU</b>	
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv	<b>Version:</b> 1_4 Final	4/13

# CONTENTS

<b>1. INTRODUCTION</b> .....	<b>5</b>
<b>2. ALIGNED INTERNATIONAL COLLABORATIONS AND ACTIVITIES</b> .....	<b>6</b>
<b>3. FEEDBACK ON INDIVIDUAL ODIN WORKING PACKAGES</b> .....	<b>8</b>
3.1. HUMAN AND THE SOCIAL SCIENCE (HSS) PROOF OF CONCEPT (DELIVERABLE 3.1) .....	8
3.2. HIGH-ENERGY PHYSICS (HEP) PROOF OF CONCEPT (DELIVERABLE 3.2) .....	9
3.3. CONCEPTUAL MODEL OF INTEROPERABILITY (DELIVERABLE 4.1) .....	9
3.4. GAP ANALYSIS AND ROADMAP (DELIVERABLE 5.1) .....	10
<b>4. SUMMARY</b> .....	<b>12</b>
<b>5. BIBLIOGRAPHY</b> .....	<b>12</b>



	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final

## 1. INTRODUCTION

The rapid development of the research infrastructure in recent years has encouraged a vast number of collaborative and cross-disciplinary research projects. A large number of these projects not only rely on collaborative access to research data, but also produce a significant number of research datasets and publications with authorships across multiple institutions. Such cross institution collaboration necessitates interoperability between research management systems and highlights the need for cohesive approach to unique identifiers and author disambiguation.

ODIN is a project funded by the European Commission to investigate identifier interoperability between two global identifier initiatives, DataCite and ORCID. The problem under analysis is global and requires transnational solutions. The ODIN project was expressly designed to have international participation in order to ensure that this European-funded interoperability initiative was in fact globally relevant, supported by global infrastructures, and informed by diverse international use cases. This report – Work Package 6 – was undertaken to validate the ODIN approach from an international perspective, thus providing a sensible risk management plan to guard against parochial approaches.

The goal of the ODIN project is to address the issue of interoperability by investigating and resolving issues related to persistent identifiers. As part of this project, international partners (ANDS, arXiv and Dryad) are investigating how the proposed ODIN solutions address the concerns of the international audience, and how the proposed interoperability model can be extended beyond the ODIN proof-of-concept projects.

The ODIN project international partners have already had significant input into all ODIN work packages. This report (ODIN Deliverable 6.1) discusses the following aspects of the ODIN proof-of-concept projects, proposed frameworks and roadmaps (WP3, WP4 and WP5):

- The broad relevance of the ODIN findings and approach and how they relate in particular to the Australian and US research sectors, as well as other international activities which are aligned with the ODIN deliverables.
- Consideration of the international context of D3.1, D3.2, D4.1 and D5.1 based on the experience on Australian and US partners.

Some of ODIN's achievements such as the roadmap for identifiers, interoperability challenges, aggregation of metadata descriptions and trust have been enabled by the efforts across multiple work packages. These achievements are not only important to ODIN partners but have a significant value to the wider community who work on research data infrastructure. In the following sections of this report we will discuss some

	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final 6/13

of the national and international activities that are aligned with ODIN activities, and we will describe the achievements from the individual work packages.

## 2. ALIGNED INTERNATIONAL COLLABORATIONS AND ACTIVITIES

The recent statement<sup>1</sup> (June 2013) by the G8 Science Ministers recognises the need for the international collaborative efforts to enable the global research infrastructure. Aligned with this statement, the ODIN project has contributed to the international collaborative efforts for interoperability among research infrastructures. Collaboration with international ODIN partners has fostered new activities that contribute to better discovery and reuse of research data, and reinforces actions taken by many different stakeholders internationally:

- **Identity Awareness of Research Data:** The initial gap analysis by ODIN partners fostered a new project by ANDS to investigate the connections in the research ecosystems that enable discovery and reuse of research data (Aryani & Burton, 2012) and a study of identity-awareness and reuse of research data in the area of social computing (Hayes, et al., 2013).
- **Research Data Alliance (RDA):** The ODIN approach on identifiers is already having impact within the following working groups (WG) in Research Data Alliance<sup>2</sup>, a joint venture between the European Commission, the Australian Government and the US National Science Foundation:
  - *PID Information Types WG:* This working group focuses on cross-community concerns regarding persistent identifiers as the core of proper data management and access, informed directly by the outputs of ODIN WP4.
  - *Data Description Registries Interoperability WG:* At time of submission of this report a new RDA working group is being launched to promote interoperability of Data Description Registries, enabled by ODIN's effort on creating a roadmap and gap analysis for interoperability between identifiers (WP5).
  - *Publishing Data Interest Group:* This group brings together all stakeholders involved in publishing research data including researchers, discipline specific and institutional data repositories, academic publishers, funders and service providers. The goal is to address the implementation of workflows for publishing data and therefore help establish appropriate supporting infrastructure.

<sup>1</sup> <https://www.gov.uk/government/news/g8-science-ministers-statement>

<sup>2</sup> <https://rd-alliance.org>

	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final   7/13

- **DataCite:** ODIN's approach is to adopt existing national schemes that can be interlinked with a coherent global framework, with each nation's local knowledge and strongly-grounded infrastructures serving as the sustainable substrate of the global scheme. This is realized by DataCite's structure, with members in 14 countries, each serving a local network of members. For example, the time of writing, 23 Australian research institutions are taking advantage of ANDS services to assign DOIs to datasets through DataCite. Dryad mints DOIs through the EZID<sup>3</sup> service at the California Digital Library, which currently has dozens of other clients, mostly in the US. arXiv similarly has plans to assign DOIs for datasets.
- **ORCID:** ORCID is a strongly international organization, with over 80 members from a variety of different stakeholder groups. ANDS has integrated ORCIDs into its researcher identity system and there are a number of ANDS-ORCID activities planned for 2013-14 including joint webinars, community building, policy development, and systems integration (Burton & Aryani, 2013). Dryad recently announced plans for integration of ORCIDs into DSpace, the widely used open source institutional repository platform (Haak & Vision, 2013).
- **Data Citation Tracking:** The ODIN ecosystem outlined in the D5.1 Gap Analysis and Draft Roadmap depicts a global open interoperable identifier system as a powerful enabling infrastructure for third party services, products, and innovation. For example, the presence of a trusted DOI registrar for data has spurred the development of the new Thomson Reuters Data Citation Index, which takes advantage of the accessible and standardized metadata from the wide variety of different DataCite clients, including Research Data Australia (by ANDS) and Dryad. Thomson Reuters has also integrated ORCIDs into their ResearcherID system, making possible in the future better tools for linking of researchers and their research data.
- **Research Funders:** A number of funders, both public and private, are members of ORCID, including Wellcome Trust, the UK National Institute for Health Research, and funders outside Europe such as the US National Institutes of Health, the US Department of Energy and the Japan Science and Technology Agency. ANDS has initiated a discussion with Australian research funders to incorporate the ODIN approach on researcher identifiers by using ORCID as the identifier for the applicants and grantees. In the US, NSF has modified its biosketch format for applicants, allowing data to be listed among articles as of relevance for evaluation, while the NIH has directly integrated ORCID identifiers into its new SciENcv system<sup>4</sup>. In the UK, a broad group of sector bodies and funders, including HESA, HEFCE, RCUK, the Wellcome Trust, ARMA, UCISA and Jisc have signed a joint statement expressing their

<sup>3</sup> <http://n2t.net/ezid>

<sup>4</sup> [http://www.nlm.nih.gov/pubs/techbull/so13/so13\\_sciencv.html](http://www.nlm.nih.gov/pubs/techbull/so13/so13_sciencv.html)



	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final 8/13

support for the ORCID initiative. This group has now been joined by RLUK and SCONUL<sup>5</sup>.

- Research Institutions:** Research institutions are beginning to incorporate persistent identifiers for researchers and data into their systems and processes. ORCID received funding from the Sloan Foundation to competitively award support for adoption and integration to nine different organizations in the US, not just for reporting but also researcher-facing data platforms such as nanoHub and Reactome<sup>6</sup>. ANDS has worked with research organisations in Australia to promote awareness of the ODIN approach to allocating ORCID identifiers to researchers who create data as an output of their project. Technical integration and interoperability with ANDS systems is under development (see section 3.3) and broad community support activities have been initiated. Under arXiv's recent institutional membership model, member institutions want better statistics about contributions from their faculty. arXiv is promoting the use of ORCID iDs, and the affiliation information that will be associated with them, as an international and general solution to this problem, rather than adopting an arXiv specific approach. These activities are aligned with and informed by the ODIN WP5 roadmap and gap analysis which highlights the need for cultural change, support and training, as well as easy to use researcher facing tools that remove the complexity of the underlying systems.

### 3. FEEDBACK ON INDIVIDUAL ODIN WORKING PACKAGES

This section presents the feedback of the Australian and US partners on the ODIN deliverables D3.1, D3.2, D4.1 and D5.1. The discussion for each deliverable is consolidated from feedback provided by ANDS, arXiv and Dryad.

#### 3.1. Human and the Social Science (HSS) Proof of Concept (Deliverable 3.1)

This proof of concept study focuses on adopting identifiers for British Birth Cohort Studies, longitudinal studies that are funded by the UK's Economic and Social Research Council (ESRC) and Medical Research Council (MRC). The D3.1 report describes the state of persistent identifier adaptation in the area of humanities and social science; in addition, it outlines the preliminary models and workflows that will be explored and developed upon in year two of ODIN.

As part of the ODIN project, Australian Data Archive has provided insights into the global social science data archive landscape, especially with regard to commonalities between

<sup>5</sup> <http://www.jisc.ac.uk/publications/generalpublications/2013/future-of-orcid-in-uk-he.aspx>

<sup>6</sup> <http://orcid.org/blog/2013/09/27/announcing-orcid-adoption-integration-program-awardees>

	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final

data providers across the world. This international perspective was enhanced with conversations with data providers such as GESIS in Germany and ICPSR in the USA.

### **3.2. High-Energy Physics (HEP) Proof of Concept (Deliverable 3.2)**

The D3.2 report -- first year report on the HEP proof of concept -- investigates the challenges and the need for identifiers in the domain of high energy physics. Particularly the report focuses on providing preliminary models and workflows for meta data exchange in this domain. This information will support the activities in the second year of the ODIN project for this community.

The report highlights two key challenges in this domain: managing authorship, and aggregating information from multiple datacentres. The high number of authors on some papers within this domain (e.g papers with more than 3,000 authors) and the fact that the information about publications and underlying datasets are derived from multiple data centres pose difficult challenges for author disambiguation.

The report describes significant work in understanding data modelling and mapping between different services in the HEP scholarly communication ecosystem. While developing an extensible model, the report focuses on concrete workflows sharing data between INSPIRE and the arXiv and HepData services. Ad-hoc methods have been used to prototype the workflows using existing interfaces at arXiv and HepData, but the analysis provides guidance for improvement of interfaces and data structures to improve the interoperability of these services in year two of ODIN.

### **3.3. Conceptual Model of Interoperability (Deliverable 4.1)**

The D4.1 report focuses on the four main challenges of persistent identifiers by ORCID and DataCite including accessibility, discovery, interoperability and sustainability. The report presents the state of the art identification systems, and proposed a conceptual model to solve the interoperability between different identifiers for data and people. The proposed model uses a trusted identifier layer to mint and manage identifiers for data and authors. The model is exemplified by the use of ORCID and DataCite services to mint and link researcher identifiers and Digital Object Identifiers (DOI) for data.

The proposed model by ODIN is aligned with the existing infrastructure in ANDS services and Research Data Australia. Since 2011 ANDS has provided support for minting DataCite DOIs. More recently, ANDS implemented ORCID integration (Burton & Aryani, 2013), enabling Australian researchers to use the trusted identifier layer to link and disseminate their datasets at the international level.

	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level:</b> PU
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final 10/13

arXiv has long maintained authority records tied to author accounts and in 2008 introduced local author identifiers. Following the model proposed by ODIN, arXiv plans to integrate with ORCID to connect local author identities with ORCID iDs for interoperability. arXiv will also create DataCite DOIs for ancillary data files using the EZID service, and these data files will be linked to the ORCID iDs of their authors.

Dryad has minted DataCite DOIs for its data since 2010 but has so far lacked the ability to recognize when authors on different datasets correspond to the same person. ODIN has helped inform development work now underway to allow user accounts to be linked to ORCID IDs, to enable coauthors editing privileges to datasets via their ORCID identities, and to publicly index ORCIDs for data authors in the DataCite Metadata Store.

### 3.4. Gap Analysis and Roadmap (Deliverable 5.1)

The D5.1 report focuses on the roadmap toward enabling a global and open discoverability of data and other scholarly materials. The report highlights the need for a global interoperable data exchange framework, and highlights the role of identifiers in linking researchers to their work and encouraging data citation and data sharing.

The report identifies four key stakeholders who are heavily involved in the design, implementation and adoption of the persistent identifiers including e-infrastructure operators, scientists, funders and scholarly communication community (publishers and libraries). Aligned with the ODIN vision, ANDS has been liaising with these stakeholders to establish persistent, transparent and sustainable connections among research data, publications, researchers and research grants. In addition, ANDS has been actively working with the community to make the research data connected, reusable, managed and discoverable. The following table shows the international perspective to the ODIN objectives and the issues identified by the gap analysis report.

ODIN Gap Analysis	International perspective
Gap 1: There is only limited access to e-infrastructures for small institutions or interest groups.	Universally relevant: <ul style="list-style-type: none"> <li>• Long tail of research is not necessarily catered for by national scale infrastructure</li> <li>• Disciplinary repositories such as arXiv &amp; Dryad may provide support to all members of a community</li> </ul>

	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final 11/13

Gap 2: There are research communities that have little or no experience with interoperable persistent identifiers.	<p>Shared international concern:</p> <ul style="list-style-type: none"> <li>• RDA working group formed to help address this issue</li> <li>• Awareness raising and support is an increasing priority, e.g. in the Australian research infrastructure strategy (National Collaborative Research Infrastructure Strategy) and the ORCID Adoption and Integration Program</li> </ul>
Gap 3: There are research communities that have built local PI systems.	Widespread phenomenon
Gap 4: There is a lack of support and funding to implement international standards.	Significant obstacle worldwide, but at least partly overcome by organizations such as ORCID and DataCite that can leverage resources in many different countries and have distributed long-term revenue streams.
Gap 5: Methods and tools to track re-use of research data and other scholarly materials are lacking.	Critical for global information systems. Sufficient public infrastructure investment can help spur the development of both non-profit (e.g. ImpactStory.org) and commercial (e.g. Thomson Reuters Data Citation Index) infrastructure.
Gap 6: Policies to encourage data sharing and acknowledge re-use in the grant review process are nascent.	<p>On going transnational concern</p> <ul style="list-style-type: none"> <li>• US National Science Foundation and National Institutes of Health have made significant first steps in this regard.</li> </ul>
Gap 7: Reliable discovery services for research data and non text-based scholarly materials are missing.	<p>Significant international lacuna</p> <ul style="list-style-type: none"> <li>• RDA working group formed through ODIN leadership</li> </ul>

	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>		<b>Dissemination level: PU</b>
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv		<b>Version:</b> 1_4 Final 12/13

Gap 8: Incentives for making datasets re-usable are unclear or missing.	One of the most powerful incentives will be the ability to track re-use of data, so (as per gap 5) sufficient public infrastructure investment can help spur the development of both non-profit and commercial infrastructure.
Gap 9: Value-added services that can incentivize citation and open science require an interoperable PI infrastructure.	Important international opportunity, with considerable experimentation underway in non-profit and private sectors <ul style="list-style-type: none"> <li>e.g. Mendeley, Dryad, FigShare</li> </ul>
Gap 10: Unique attribution and linking between researchers, their scholarly materials and funding will only be possible with the collaborative adoption of global and interoperable PI systems.	Underlying global concern <ul style="list-style-type: none"> <li>Experience within the very successful collaborations between INSPIRE, arXiv, and ANDS has highlighted the need for interoperable PIs to match identities between systems.</li> </ul>

#### 4. SUMMARY

In this report, we have identified the relevance of the ODIN 1st year outcomes to the international audience, and we have demonstrated a number international activities aligned with ODIN findings. The scope of these activities highlights the importance and versatility of ODIN interoperability model, and suggests that presented proof of concept projects can be further developed to benefit extended community. In addition, the gap analysis by ODIN partners provided a cohesive list of requirements for future development of e-infrastructure and interoperability projects in the research sector.

#### 5. BIBLIOGRAPHY

- Aryani, A., & Burton, A. (2012). Identity Awareness: Toward an Invisible e-Infrastructure for Identifying Data and Authors. *eResearch Australasia*. Sydney: DOI: 10.6084/m9.figshare.155650.
- Burton, A., & Aryani, A. (2013). ORCID integration: A case study from ANDS and international development. *eResearch Australasia*. Brisbane: IEEE DOI: 110.6084/m9.figshare.799755.
- Haak, L., & Vision, T. (2013). *ORCID: Interoperability and discovery*. From OAI8 DSpace Users Meeting: <http://www.slideshare.net/ORCIDSlides/orcid-oai8-20130618>
- Hayes, A., Mann, S., Aryani, A., Sabine, S., Blackall, L., Waugh, P., et al. (2013). Identity Awareness and Re-use of Research Data in Veillance and Social Computing. *International Symposium on Technology and Society*. Toronto: IEEE DOI:

 ODIN	<b>D6.1 International validation of D3.1, D3.2, D4.1 and D5.1</b>		
	<b>WP6 Internationalisation</b>	<b>Dissemination level: PU</b>	
	<b>Authors:</b> Australian Data Archive, DRYAD, arXiv	<b>Version:</b> 1_4 Final	13/13

10.1109/ISTAS.2013.6613101.

