

Grant agreement no. 312788



<http://odin-project.eu>

D5.1 Gap analysis and draft roadmap

WP5 – Strategy

V1_0

Final

ODIN is co-funded by the EC under the e-Infrastructures Activity of the FP7 Capacities Specific Programme.

© Copyright 2013 ODIN Consortium. Some rights reserved. This work is licensed to the public under the Creative Commons Attribution 3.0 License. <http://creativecommons.org/licenses/by/3.0/>



	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Abstract: During the first year of ODIN several steps have been taken to study the strengths and weaknesses in PIDs across disciplines and stakeholder groups. Particular emphasis has been given to the opportunities ahead – the potential that can be unlocked by an open and interoperable PID framework in Europe and beyond.

To do so, a continuous exchange with experts and users of PIDs has been in the focus during this first year. This has happened during selected events, as well as dedicated interviews and literature review. As the core result a SWOT analysis is presented, as well as a concise gap analysis. The report concludes with recommended actions to address the main gaps.

The analysis highlights that the gaps must be addressed urgently and mainly in a collaborative manner. An open, global and interoperable PID layer could unlock the full potential of Open Science if all stakeholder groups and disciplines are involved.



Lead beneficiary: CERN

Date: 18/09/2013

Nature: Report

Dissemination level: PU (Public)

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final 3/57

Document Information

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

Deliverable	Number	5.1	Title	Gap analysis and draft roadmap
Work package	Number	5	Title	Strategy

Document identifier	ODIN-WP5-Gap-Analysis-Draft-Roadmap-0001-0_0		
Delivery date	Contractual	Month 11	Actual 12
Status	Version 1_0		Final <input checked="" type="checkbox"/> Draft <input type="checkbox"/>
Nature	Report <input checked="" type="checkbox"/> Prototype <input type="checkbox"/> Demonstrator <input type="checkbox"/> Other <input type="checkbox"/>		
Dissemination Level	<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)		

Authors (Partner)	Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		
Responsible Author	Sünje Dallmeier-Tiessen	Email	Sunje.dallmeier-tiessen@cern.ch
	Partner	CERN	Phone +41227674103

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Document Status Sheet

Issue	Date	Comment	Author
0.1	15 th July 2013	First draft	CERN
0.2	15 th Aug 2013	Updated draft after CERN internal review	CERN
0.3	27 th Aug 2013	Restructured draft	CERN
0.4	4 th Sep 2013	Draft for project review	CERN
0.5	9 th Sep 2013	Updated version after review by Laure Haak (ORCID)	CERN, Laure Haak (ORCID)
0.6	13 th Sep 2013	Updated version after review by Amir Aryani and Adrian Burton (ANDS)	CERN, Amir Aryani (ANDS), Adrian Burton (ANDS)
0.7	16 th Sep 2013	Updated version after review by Salvatore Mele (CERN)	CERN
1.0	18 th Sep 2013	Final version	CERN

Document Change Record

Issue	Item	Reason for Change

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

CONTENT

1. INTRODUCTION	6
2. APPROACH AND CHALLENGES DURING THE FIRST YEAR OF ODIN	10
3. RELATED STUDIES	12
3.1. DIGOIDUNA.....	13
3.2. APARSEN	14
3.3. KNOWLEDGE EXCHANGE: THE DEN HAAG MANIFESTO	15
4. ODIN KICK-OFF EVENT	15
5. SYNERGIES WITH OTHER ODIN WORK PACKAGES: DISCIPLINARY PROOFS OF CONCEPT AND INFRASTRUCTURAL INTEROPERABILITY	20
6. RESULTS FROM THE FIRST YEAR (SWOT ANALYSIS)	21
6.1. STAKEHOLDER SPECIFIC SWOTs IDENTIFIED DURING THE FIRST YEAR OF ODIN	21
6.2. SUMMARY OF SWOT ANALYSIS	30
7. GAPS	33
8. FROM THE GAP ANALYSIS TO A DRAFT ROADMAP	35
9. RECOMMENDATIONS FOR FUTURE ACTIONS: DRAFT ROADMAP	38
10. CONSIDERATIONS FOR THE FUTURE ACTIONS AND ODIN 2ND YEAR	45
11. SUMMARY	48
12. APPENDIX 1: INTERVIEWS	49
12.1. SHEET FOR THE INTERVIEWEES TO COMMENT ON.....	49
12.2. QUESTIONNAIRE.....	54
13. APPENDIX 2: LIST OF ABBREVIATIONS	57

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

1. INTRODUCTION

Contemporary research is being transformed by a data deluge. In some scientific disciplines, data production has risen steeply in the past decades due to faster and better instruments and techniques. In disciplines such as cultural heritage, more data are being digitized and made available publicly. In the social sciences and humanities, the Opportunities for Data Exchange FP7 project (ODE)¹, in which ODIN partners have been involved, has shown that data sharing is becoming more prevalent with the advent of “digital humanities”. In all domains of research, data is an increasingly significant input and output of the research process. With practices in flux and the production and significance of data on the increase, it is now crucial to establish a framework for sharing and attributing research data. This framework requires tools, infrastructure and documentation: tools to allow researchers to better and more easily share, discover, and access research data; infrastructure and documentation to allow communities to make use of such data, now and into the future.

For such a framework to be successful, we need new incentive mechanisms for researchers, such as amended research assessment practices which accommodate open sharing of digital scholarly objects including data. Tools must be embedded in a framework in which access and openness are at the core. The challenge is bringing together in this framework many interconnected issues: societal, policy, technical, and business, while building trust between data generators and user communities. The

¹ Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Herterich, P., Lambert, S., Mele, S., Nordling, J., Pfeiffenberger, H., Ruiz, S. & Smit, E. (2012). D6.1 Summary of the Studies, Thematic Publications and Recommendations. Retrieved from http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/11/ODE-WP6-DEL-0001-1_0.pdf.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	7/57

challenge is now clear as described in the landmark EC High-Level Expert Group report "Riding the Wave"², and the results of the ODE project.

To enable the global and open discoverability of data and other scholarly materials, it is indispensable to have a global interoperable data exchange framework. Digital identifiers are fundamental to this framework. They are the key element in linking data and user communities. Digital identifiers also are fundamental for enabling incentive schemes such as data citation. This is where the ODIN project fits in. ODIN is articulating a "missing thin layer": a framework, or a series of interconnections, that support interoperability among existing Persistent Identifiers (PID) for people and data to enable access and attribution.

PIDs provide stable linking to online resources, support long term preservation of digital resources, and assist in monitoring resource usage. However, widespread adoption of PIDs is so far only well established for DOIs and particularly only for research articles published in scholarly journals. While DOIs and other PIDs are increasingly being embedded in research datasets, there remain challenges with discoverability across repositories and consequent implications for attribution and usage tracking. In the last couple of years, through a crucial acceleration, PIDs are now emerging for authors and other contributors, including ORCID identifiers or ISNIs. One of the main challenges is to understand the current state of the interoperability of these PIDs, for research data and contributors, or lack thereof. This is to ODIN the missing piece that hinders the development of an open (PID based) infrastructure supporting data-intensive science.

² <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Four key stakeholder communities are heavily involved in the design, implementation and adoption of such a PID thin layer and stand to benefit from its existence.

1. **e-Infrastructure and data centre operators** who need to serve community needs, implement and adhere to standards, and are ultimately responsible for supporting an interconnected fabric of data-intensive research and scholarship. Moreover, third-party providers of value-added services (such as impact assessment and information discovery) are increasingly emerging.
2. **Scientific and research communities** who rely on trusted e-Infrastructure to thrive in the data-intensive era, and ultimately evolve into an e-Science perspective, where research outputs are routinely shared and re-used..
3. The **scholarly communication community** including publishers and libraries who are the main actors in disseminating and discovering research, to a scientific community in flux towards data-intensive science.
4. **Funders and policy makers** who are responsible for funding research in general and e-Infrastructures in particular, with a need to increase collaboration and ultimately optimise investment of public funds.

It is crucial to understand the challenges and obstacles for an interoperable PID framework to emerge in the joint interest, and with the joint effort, of these stakeholders. The major focus of the ODIN project is to determine how PIDs for data and contributors are used within these communities, and what is further needed to develop such a framework to enable data sharing, attribution and discovery. ODIN is primarily interested in how to leverage PIDs and related digital science/scholarly communication services that exist today, while suggesting the way ahead to those being developed by one or more stakeholder groups. To this end, ODIN has engaged with stakeholders to analyse current PID practices and infrastructures while paying special attention to some smaller and niche communities without immediate access to large scale projects and funding. ODIN is particularly interested in identifying the scenarios where not using standard PIDs, the lack of interoperable

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

identifiers, or not connecting data and contributor identifiers effectively hinders data sharing and cross-disciplinary scientific opportunities.

Through these consultations, ODIN partners have identified a number of very concrete requests from all stakeholder communities:

- Knowing how to resolve an identifier for data or contributors to the original entity
- Knowing how to connect identifiers for the same object or contributor across different services
- Knowing how to parse an entire identity chain, connecting data, people and publications, starting from any of the links
- Knowing how to find all the publications that use a particular dataset
- Knowing how to find all the dataset from a particular contributor
- Knowing how to find all contributors who participate in producing, curating, and enriching a dataset
- Knowing how to answer questions like "What other papers has this author written?"
- Knowing under which licenses a dataset is released

Chapter 2 gives an overview of the approach leading to the results presented in this deliverable. Chapter 3 summarizes some relevant studies which informed ODIN. Chapters 4 and 5, respectively, describe the relation between ODIN results from the Kick-Off event (chapter 4) and other work packages (chapter 5) with the work of this report. The results of an extensive stakeholder-based SWOT analyses are presented in chapter 6. Chapter 7 presents a gap analysis for the onset of a appropriate PID framework. Chapter 8 traces the way from this gap analysis to a first draft roadmap, which is presented in Chapter 9. Chapter 10 gives an overview of considerations for future actions in ODIN year 2 followed by a short summary of the report in Chapter 11.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	10/57

2. APPROACH AND CHALLENGES DURING THE FIRST YEAR OF ODIN

This deliverable presents a first draft of a gap analysis and a roadmap. The work presented will be reviewed throughout the second year of the project, including results from the other activities in the project (from desk-research to hands-on implementations, from further consultations to open discussions) and thus will include latest relevant developments. The project members represent leading international players working with PIDs such as ORCID and DataCite as well as information platforms from different scientific communities (life sciences, high-energy physics, humanities and social sciences). The gap analysis and roadmap are informed by listening to a diverse set of researchers and infrastructure experts working with PIDs, studying best practice examples, and integrating this information with the project members' own expertise. With this continuous input from external experts, the approach ensures the timely involvement of international developments in the area (see Figure 1).

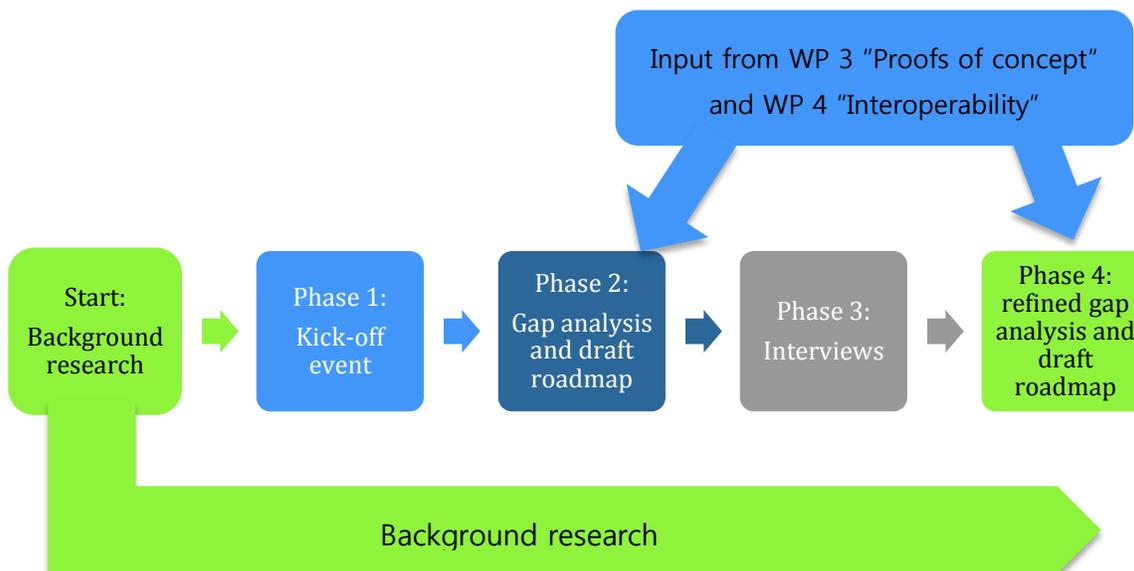


Figure 1: Visualisation of ODIN WP 5 approach

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

The work on this report started with background research and preparations for the ODIN Project Kick-off event. There, presentations by a diverse set of stakeholders provided our first external input. Literature review was conducted continuously during the project. In parallel, the strong collaboration between ODIN partners and work packages provided in depth input on technical and disciplinary aspects of PIDs. These internal discussions were consolidated and incorporated into a dedicated section in this report. Both internal and external consultations fed into the development of this draft of the gap analysis and roadmap (phase 2). We also developed a two page summary of input and findings to discuss with external experts in another round of consultation during phase 3.

We conducted interviews with 10 PID experts or users, selected based on their discipline, field of expertise, and role (e-Infrastructure provider, researcher, librarian, publisher, funders). The interviewees were from the following areas: economy and humanities and social sciences, biomolecular and medicine, and geosciences. They were e-Infrastructure providers (e.g. large scale repositories for publications and data, third-party services in e-science), librarians (national libraries to research libraries), publishers (from HSS and science domains) and funders (for the science domain). The materials from the interviews are given in the appendix. Most of the interviews were conducted in Oxford, during the ORCID/Dryad conference in May 2013, which gathered key world experts in this field³. The rest were conducted by teleconference.⁴ The interviews helped us to validate, revise, and refine the gap analysis and recommendations for future actions in phase 4. The interviewees generally corroborated ODIN's internal gap analysis but highlighted different challenges for

³ <http://orcid.org/orcid-outreach-meeting-symposium-and-codefest-may-2013>.

⁴ Notes were taken during the interviews, but these are not included here due to the detailed specifications discussed and the impossibility of anonymizing them. They are available to project members on the Twiki.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

particular PID stakeholder groups. We addressed this in the draft by breaking out gap analysis by group in chapter 6.1.

As highlighted before, the ODIN project is very much aware that PIDs and related services are under continuous development, with engagement by a diverse set of stakeholders. The ODIN partners have been able to study PID frameworks in 'the making' as third-party entities (from pro-bono or small start-ups all the way to large conglomerates) have started to build PID-enabled services on open science materials. It has been clear that all stakeholders experience limitations in adopting and implementing PID systems as there are no comprehensive infrastructures with interoperable identifiers they can rely on. Phase 4 is described in chapters 6 to 9 of this document, starting with the SWOT analyses in chapter 6 and the gap analysis in chapter 7. Chapter 8 summarizes the way from the gap analysis to the draft roadmap which is then presented in chapter 9.

3. RELATED STUDIES

The landscape of PID systems is changing rapidly, but is still fragmented. While the number of candidate PID infrastructures has decreased, there is still no single solution that offers all required features. Studies such as DIGOIDUNA⁵ have set important precedents in exploring the value of PID systems and associated implementation challenges. We found an increasing awareness among stakeholders - including funders and policy makers - of the value of PID systems for supporting research data attribution and usage monitoring. This chapter will summarize three important studies on PID infrastructures (DIGOIDUNA, APARSEN and the Knowledge Exchange Den Haag Manifesto) as the main outcome of the background research on which the results presented in this document are based.

⁵ Digital Object Identifiers and Unique Authors Identifiers to enable services for data quality assessment, provenance and access (DIGOIDUNA), <http://www.digoiduna.eu/>.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

3.1. DIGOIDUNA

On behalf of the European Commission, the DIGOIDUNA team studied "Digital Object Identifiers and Unique Authors Identifiers to Enable Services for Data Quality Assessment, Provenance and Access"⁶ in 2011. The project report consists of an analysis of the role of PID in e-Infrastructures, a strengths, weaknesses, opportunities and threats (SWOT) analysis of the current situation regarding PIDs, and recommendations on how to develop an open and sustainable e-Infrastructure for persistent object and author identifiers. ODIN is closely collaborating with the DIGOIDUNA team to build on these findings.

The report identifies that an increasingly interdisciplinary and international research community needs an e-Infrastructure that supports data access and sharing across national, organizational, disciplinary, cultural and technological boundaries. Europe has made significant investments in e-Science in the last few years. An interoperable PID infrastructure is a crucial element to maximise the return on this investment. The technical components for implementing such an infrastructure are available. In addition to local ad hoc solutions, there are initiatives such as DOI, Handle, and ORCID that approach the topic from a cross-domain perspective. However, a remaining challenge is the lack of consensus and coordination among parties. Solutions and developments are driven by the needs of a few active stakeholders, rather than common policies on PID governance. Furthermore, financial sustainability is still an unsolved issue, as traditional funding schemes seem to be not sufficiently flexible or diverse to provide sustainable financing. Last, the success of PIDs is dependent on their use by the research community. Without the will to develop and adopt community practices, no common goals can be reached.

⁶ Bouquet, P., Bazzanella, B., Riestra, R., & Dow, M. (2011). DIGOIDUNA: Digital Object Identifiers and Unique Author Identifiers to enable services for data quality assessment, provenance and access. Tratto da DIGOIDUNA, Digital Object Identifiers and Author Identifiers: http://digoiduna.files.wordpress.com/2011/12/digoiduna_final_report_expert_feedback1.pdf.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Based on the SWOT analysis, DIGOIDUNA suggested four major actions to create a coordinated and working PID infrastructure:

1. A common agenda for a governance model and an integrated infrastructure for managing PIDs should be defined.
2. The implementation of the defined PID agenda should be bootstrapped to ensure a certain amount of coordinated PID systems.
3. Already active stakeholders should promote interoperability and established PID systems to attract more users and prevent the creation of non-interoperable ad hoc solutions.
4. A sustainable business model for PID systems should be developed

3.2. APARSEN

The EU project APARSEN has surveyed⁷ 103 primarily European research institutions, on the usage of PIDs. The APARSEN survey results show that cross-disciplinarity, trans-nationality and management by trusted institutions are the most important functional requirements. In addition, respondents stated that PID systems should provide citability and a global resolution service. Almost 90% of the responding institutions use PIDs such as DOIs, Handle, or URN to identify their digital objects. However, they report low internal adoption of the PID service. About half of the services analysed included author identifiers, 21% of the institutions use their own ID system. 25% of the surveyed institutions plan to adopt ORCID identifiers in the future. Barriers to PID adoption included both a lack of understanding of the benefits of use coupled with a lack of knowledge of their existence. Based on these findings, APARSEN has been working to identify a common data exchange standard to

⁷ http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D22_1-01-1_9.pdf.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

support a query tool that works across and leverages existing fragmented PID systems.

3.3. Knowledge Exchange: The Den Haag Manifesto

The Den Haag Manifesto⁸ came out of a seminar on Persistent Object Identifiers organised by Knowledge Exchange in June 2011. The manifesto states six principles intended to be a basis for a co-ordinated approach to identifier issues across the PID and Linked (Open) Data (LOD) communities:

1. Ensure PIDs can be referred to as HTTP URI's, including support for content negotiation.
2. Use appropriate LOD vocabularies to populate schema elements.
3. Identify the minimum common set of schema elements across different kinds of identifiers in the scholarly communication space.
4. Use same-as relations to help PID interoperability across PID systems/schemas
5. Use PIDs for Subjects and Objects
6. Work with the LOD community on simple policies/procedures to improve persistence of HTTP URIs

4. ODIN KICK-OFF EVENT

ODIN hosted a meeting to kick-off the project, in October 2012. The event was attended by the stakeholder groups involved in PID management, from a range of disciplines and cross disciplinary services, like CrossRef. The individual presentations highlighted the experience and progress within the individual services presented.

⁸ <http://www.knowledge-exchange.info/Default.aspx?ID=462>.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

The meeting aimed to discuss each of the key original challenges in PIDs underpinning e-Science, as identified by ODIN. These are Access, Discovery, Interoperability, and Sustainability:

Access: Researchers must be able to seamlessly access datasets used in a journal article, a grant report, or another scholarly artifact, with explicit and unambiguous terms of use

Discovery: Researchers must be able to identify scholarly works or datasets related to each other either through networks of references and citation or through shared contributors

Interoperability: Users must be able to connect datasets and contributors across independent platforms using different, but interoperable identifier schemes.

Sustainability: Funding agencies and research institutions must be able to depend on the ongoing viability of the systems that provide and maintain persistent identifiers.

Each of the challenges has a corresponding objective for ODIN, as a guide for the project. The input gathered during the kick-off event validated these objectives and provided clear and concrete plan for actions. All details are provided in ODIN Deliverable D2.1, and summarised in Table 1. This approach ensured the applicability of ODIN results to all stakeholder groups and disciplines from the very beginning without losing attention to the detail. The concrete actions are an important guideline for the process described in this Deliverable, and have informed the consultations leading to the following results.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

In addition to defining specific actions, kick-off meeting participants reached consensus on a set of key questions at the core of the international and interdisciplinary conversation on PIDs for data and contributors. These questions have provided a central structure for ODIN and have guided the research and interviews underlying this gap analysis.

- Are persistent identifiers for data and contributors the goal or rather tools to solve specific problems?
- Is there sufficient communication among the different stakeholder communities?
- How do we most efficiently promote data citation?
- How do challenges differ for small and large datasets?
- How to accommodate local differences while not losing sight of global commonalities?

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final 18/57

Table 1 Overview of objectives, activities, outputs, and actions for the main ODIN challenges

Challenge	Objective	Activity	Output	Actions, from the kick-off meeting
<i>Access</i>	Build shared understanding	Document perceptions, attitudes and barriers	Roadmap of outstanding issues and recommendations	Encourage discussion and collaboration between communities. ODIN partners will engage with the research community to obtain feedback on the draft gap analysis and roadmap. The kick-off meeting highlighted a need to get insights beyond the WP3 proof of concept HEP and HSS.
<i>Discovery</i>	Navigation across data and contributors	Define and implement workflow scenarios for attribution in Humanities and Social Science (HSS) and High Energy Physics (HEP)	Proofs of concept at two disciplinary extremes	Address specific issues of working with many small heterogeneous datasets. A first step is to facilitate the navigation between an author/contributor and their data. ODIN partners also need to understand how an interoperable PID layer can support more complex cases, e.g. multiple datasets, multiple versions, multiple publication outlets.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	19/57

Challenge	Objective	Activity	Output	Actions, from the kick-off meeting
<i>Interoperability</i>	Demonstrate interoperability	Define mappings between selected identifier systems of DataCite and ORCID members	Mechanisms for integrating multiple persistent identifier systems	Identify main interoperability use cases and define their business value. ODIN focuses on two PID systems, ORCID and DataCite. We must expand include other PID systems, such as ISNI and CrossRef and understand how these systems must interoperate (if at all) to meet community needs.
<i>Sustainability</i>	Stimulate value-added services	Build developer community to exploit interoperable persistent identifiers	Model of interoperable and open identifiers validated at a CodeFest	Encourage development of novel tools to track the reuse of data. ODIN should build on its wide partnership and encourage grassroots software development that builds on an open and interoperable PID layer, including tools that leverage ORCID and DataCite APIs.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

5. SYNERGIES WITH OTHER ODIN WORK PACKAGES: DISCIPLINARY PROOFS OF CONCEPT AND INFRASTRUCTURAL INTEROPERABILITY

This gap analysis draws upon other parallel ODIN Work Packages (WP), especially WP 3 “Proofs of concept” and WP 4 “Interoperability”.

WP3 develops proofs of concept for open and interoperable PIDs for data and contributors in Humanities and Social Sciences (HSS) and High Energy Physics (HEP). The work accomplished within ODIN WP3 allows for a first assessment of the similarities and differences between HSS and HEP. While we plan to expand this analysis to other disciplines, already we have been able to determine that open and interoperable PID standards, such as DOI and ORCID iDs are being established and connected to previously local/community identifiers. HSS and HEP communities provide useful examples on how PIDs can be integrated into a wider interoperable framework.

ODIN research in WP4 investigates scenarios for building interoperable infrastructures that leverage data exchange standards and PIDs for data and contributors. The open, discipline-neutral and inclusive conceptual model developed in this work package consists of:

- The trusted identifier layer, which includes criteria for PIDs for objects and people
- The data citation virtuous circle linking research data and contributors
- Common data services e-Infrastructures that provide linked and persistent identifiers for the European e-Infrastructure framework.

Interactions with the HSS and HEP communities and key PID infrastructure providers, ORCID and DataCite, allowed us to identify important questions for the expert interviews and have provided a practical context for the gap analysis and draft roadmap.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

6. RESULTS FROM THE FIRST YEAR (SWOT ANALYSIS)

The step by step approach of ODIN comprises a first SWOT analysis (phase 2), interviews and feedback (phase 3), leading to a revised and refined set of recommendations (phase 4). The gap analysis highlights where and how the “missing thin layer” limits the impact of existing PID services and identifies actions needed to close these gaps. The analysis is split in two parts: the first (described in section 6.1) is a SWOT analysis focused on individual stakeholder groups; the second (described in section 6.2), identifies and discusses common aspects across stakeholders.

6.1. Stakeholder specific SWOTs identified during the first year of ODIN

6.1.1. e-Infrastructure providers and data centres

e-Infrastructure providers and PID management services are not a homogenous group. Each provider and service has emerged from a concrete disciplinary, national, or business need. Therefore, success for each is based on flexibility, establishing trust, and a clear focus on their user community.

One example of this is CrossRef⁹. A non-profit organization serving the publisher community, CrossRef receives metadata from publishers and provides services to resolve DOIs persistently for published works including journal articles and datasets. While CrossRef does not provide object-level metrics, their services can be used to support the derivation of such metrics from DOI resolutions.¹⁰

⁹ <http://www.crossref.org/>.

¹⁰ Fenner, M. Metrics and Attribution. May, 2013. <http://blogs.plos.org/mfenner/2013/05/21/metrics-and-attribution-my-thoughts-for-the-panel-at-the-orcid-dryad-symposium-on-research-attribution/>.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Data centres and repositories often serve the needs of a specific discipline or specialised subfield. The infrastructure they provide is tailored to the material they store, which, in addition to datasets of all sizes and complexity, can include analysis tools and software. Even so, there are gaps and challenges shared across providers. One is the establishment and adoption of data exchange standards to support interoperability across providers and disciplines. Another goal is the creation of global services that support attribution and citation between datasets and other object types such as journal articles. Data centres have an important role to play in facilitating data citation and have much to gain from citation counts.

ODIN WP3 results show that HSS data archives are currently working on community standards. Their processes are usually broadly based on the Open Archival Information System (OAIS). The Data Documentation Initiative (DDI) is working on creating an international standard for describing data from the field of social, behavioural, and economic sciences [for more details about HSS and examples see ODIN D3.1 HSS Proof of Concept¹¹].

Also of interest to ODIN are third-party service providers building on PID infrastructures. They offer services that support other stakeholders, in particular researchers, and can be an important component in adoption of PIDs by the community. For example, the metrics provided by Impact Story¹² or Altmetric¹³ can be incentives for researchers to share their data and for PID providers to work together to ensure compatible data exchange standards.

¹¹ Delivered to the EC on 07/08/2013, not published yet

¹² <http://impactstory.org/>

¹³ <http://www.altmetric.com/>

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Table 2 SWOT analysis for e-Infrastructure providers

<i>STRENGTHS</i>	<i>WEAKNESSES</i>	<i>OPPORTUNITIES</i>	<i>THREATS</i>
<ul style="list-style-type: none"> • <i>Many e-Infrastructures are already implementing PIDs for data and contributors;</i> • <i>Some third-party services already exist, fostering further adoption</i> 	<ul style="list-style-type: none"> • <i>Tendency to use home-grown solutions;</i> • <i>Lack of interoperability between PIDs hinders adoption and reduces benefits to users</i> 	<ul style="list-style-type: none"> • <i>Build bridges between community-grown PID infrastructures;</i> • <i>Allow data exchange/access with interoperable frameworks</i> 	<ul style="list-style-type: none"> • <i>Services that are not part of a global or cross-disciplinary layer became silos and threaten wide adoption of PIDs (and third-party services)</i>

6.1.2. Researchers

Researchers want to build on prior work and communicate their results. In addition to producing research outputs, researchers are linking to them in many ways, via citations, bookmarking, and, increasingly, discussions in social (professional) media. To support these activities, research e-Infrastructures must provide long-term trusted services that allow persistent connectivity. An important question for ODIN is how to capture links between researchers and outputs such as datasets, where the researcher is the information provider. ODIN also needs to consider how to encourage data sharing in those communities where research data is not being shared and is not accessible for others to reuse. As mentioned before, incentive mechanisms need to be put in place to engage researchers, better tracking of the research output and their assessment for example. It is important to engage research communities and their infrastructure providers to understand how PIDs can help to make research data sharing and re-use (tracking) easier and more reliable.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	24/57

Data sharing best practices are emerging in some communities. In the Earth Sciences, there is the data repository PANGAEA¹⁴. The usage of DOIs for data citation is becoming part of the publishing culture, an established and trusted workflow supported by researchers and publishers. In other research communities, PIDs have been created for specific data types. One example is in genetics, where GenBank¹⁵ identifiers for genetic sequence data are widely used and facilitate global cooperation. The challenge for ODIN and the community is to determine whether these solutions can be adapted more generally and also how these community specific data PID systems interoperate with person PID solutions. Special attention is required to understand the specific disciplinary needs, and in particular for access to sensitive datasets. Support mechanisms and workflows need to be discussed so that also such datasets can be uniquely identified and cited although they might not be accessible for everyone. Through their involvement with journals e.g. as reviewers, researchers can influence the implementation of policies for PID usage.

It is especially important that this PID infrastructure is integrated as seamless as possible in the research lifecycle. That way participating does not require extra time commitment from the researcher without receiving any benefit. The PID infrastructure should facilitate open science practices, and allow for its integration into the research reward and assessment schemes.

¹⁴ <http://www.pangaea.de/>

¹⁵ <http://www.ncbi.nlm.nih.gov/genbank/>

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Table 3: SWOT analysis for researchers

<i>STRENGTHS</i>	<i>WEAKNESSES</i>	<i>OPPORTUNITIES</i>	<i>THREATS</i>
<ul style="list-style-type: none"> • <i>Researchers are motivated by discoverability—of their own work and that of others that helps support their research. They embrace PIDs solutions that support discoverability and facilitates their work to be cited and re-used</i> 	<ul style="list-style-type: none"> • <i>The lack of an interoperable approach to attribution, citation, and measurement of re-use of data reduces the incentives for researchers to engage with PID initiatives</i> 	<ul style="list-style-type: none"> • <i>Best practices for PID systems supporting attribution and access exist in many disciplines. These examples can be leveraged to build a globally interoperable PID infrastructure</i> 	<ul style="list-style-type: none"> • <i>Disciplinary silo solutions, if easy to use and well established for a given community, present a barrier for adoption of interoperable systems and global solutions</i>

6.1.3. Scholarly communication community

Librarians

The library and information science services manage a wealth of research information derived from many different sources, and in many cases already connected to publication metadata. A big challenge is to provide these data in an interoperable format and make it accessible and usable. As libraries are a primary resource for many researchers, they serve as a very important point of integration and adoption of PID standards. While ORCID and DataCite can supply metadata and ease management of library repositories, such services need to be linked to an incentive system for the library and its users. The “thin layer” under analysis by ODIN could facilitate the incorporation of such services - by allowing the tracking of an individual’s output for example. Such integrations should become part of the default

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

services provided by institutional repositories or libraries, ideally with support from repository software providers. Activities allowing linking to external resources in the long term are important for the overall library and information science community and their service providers. To achieve this, the library and information science community need to commit to implementing interoperable PID systems. They have been at the forefront in implementing PIDs, like ISNIs, URNs, etc. However, the interoperability between these systems is still insufficient.

Table 4: SWOT analysis for librarians

<i>STRENGTHS</i>	<i>WEAKNESSES</i>	<i>OPPORTUNITIES</i>	<i>THREATS</i>
<ul style="list-style-type: none"> • <i>Experience with linking diverse source of information, knowledge databases and standards</i> 	<ul style="list-style-type: none"> • <i>Most services often focused on text publications, missing link to datasets and other scholarly materials</i> 	<ul style="list-style-type: none"> • <i>Emerging intention to collaborate with other stakeholders to added value to each other services</i> 	<ul style="list-style-type: none"> • <i>Emerging intention to collaborate with other stakeholders to added value to each other services</i>

Publishers

Traditionally, journal articles have been at the centre of scholarly communication. They are widely referenced by DOIs from CrossRef, a well established global PID system. Since the ISI first started tracking citation counts, this metric has been the sine qua non of research impact. It thus is imperative that publications are attributed correctly to their authors. ORCID offers a solution to connect authors and their

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

output and is a way for researchers to aggregate all the material they have produced, both retrospectively and with new articles. ORCID also supports attribution of other types of research works, including datasets, through integration of ORCID iDs into dataset deposition workflows by organizations including CrossRef, figshare, ANDS, and DataCite.

Research data is getting more attention as standalone research output. In the publishing realm, data have often been shared as supplementary material, and citation restricted to the paper in which a dataset was described. This solution has limited discoverability and access to data, and has led to the proliferation of many independent data archives. While these archives have supported the storage and preservation of datasets, the issue of attribution and reuse, and linkage between the article and its corresponding dataset(s) continue.

Publishers see the need to support standardised data citation alongside of the citation of articles, and practices are emerging for collaboration between publishers and data archives, such as the interaction between Scopus¹⁶ and PANGAEA. Overall, the situation is very fluid with no clear sign of emerging standards, and still some scientific communities hesitate to reuse data or use PIDs for citing data.

¹⁶ <http://www.scopus.com/>

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Table 5: SWOT analysis for publishers

<i>STRENGTHS</i>	<i>WEAKNESSES</i>	<i>OPPORTUNITIES</i>	<i>THREATS</i>
<ul style="list-style-type: none"> • <i>Early adopters and awareness of PID for scholarly materials (CrossRef) and authors (ORCID)</i> 	<ul style="list-style-type: none"> • <i>Varying level of awareness or interest in interoperability with data and technological barriers for its integration in scholarly communication</i> 	<ul style="list-style-type: none"> • <i>Capture PID for authors and non-textual material with journal articles, build links, through interoperable infrastructures, to other services.</i> 	<ul style="list-style-type: none"> • <i>Limited capture of data PID or lack of integration with PID systems may lead to “undiscoverable”/ “dead-end” material undermining incentives for other stakeholders</i>

6.1.4. Funders and policy makers

Research funders are interested in measuring the impact of their programs, by discipline and on an international scale. They expect the funded research outputs to be easily discoverable and accessible, to be reproducible and re-usable for new research and to promote public understanding of research. PIDs for people and data—and interoperation between these systems—support these goals.

Funders promote the development and use of research administration technologies, and have provided policies and funding to this end. Funders are also starting to adopt and implement PID frameworks such as ORCID iDs and CrossRef’s FundRef in their systems, enabling persistent machine-discoverable linkages between grants, grantees, and research outputs.

National funders, as well as the European Commission, are increasingly supporting Open Science practices. This means that “openness” is demanded alongside with

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

allocated funding. The policies include datasets and data management practices. For these policies and practices to be effective, it is crucial that funders work with other stakeholders to determine best practices among and across disciplinary communities for PID infrastructures, for interoperability, for implementation, and for incentivizing adoption and use by researchers.

Table 6: SWOT analysis for funders and policy makers

<i>STRENGTHS</i>	<i>WEAKNESSES</i>	<i>OPPORTUNITIES</i>	<i>THREATS</i>
<ul style="list-style-type: none"> • <i>Fast-growing awareness and strong commitment to PID infrastructures underpinning data-intensive science</i> 	<ul style="list-style-type: none"> • <i>Interoperable PID infrastructures not yet in the focus of policies/funding</i> • <i>Ad-hoc national solutions sometimes favoured instead of global ones</i> 	<ul style="list-style-type: none"> • <i>Interoperable and global PIDs underpins effective management of science</i> • <i>Incentives from PIDs result into more open science and return on investment.</i> 	<ul style="list-style-type: none"> • <i>Lack of co-ordinated support to global, participative, PID infrastructure and thus, opting for local non-interoperable solutions</i>

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

6.2. Summary of SWOT analysis

The SWOT analysis highlighted aspects relevant to the individual stakeholder groups, but also pointed to common aspects across groups, in particular the “threat” of silo solutions (see Table 7 below). Without the perceived need for interoperable PID frameworks, silo or pioneer solutions have been emerging from the different starting points and approaches. Often individual solutions, these PID systems may not have “bridges” to other PID systems, severely diminishing their potential in supporting discoverability and thereby adoption by researchers. An important aspect to consider when looking at disciplines or organisations not yet engaged in PID management is how they find appropriate partners to build e-Infrastructures, and what importance in this process are interoperable PIDs? Our SWOT analysis indicates the need for integrated services across platforms and stakeholders to provide seamless services that support researchers and thus incentivize open science. The future pathway must be a collaborative approach - where the individual partners build on their expertise, but move beyond their original domain to contribute to and participate in an interoperable PID framework.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	31/57

Table 7: Overview of ODIN SWOT analyses for all relevant stakeholders

		STRENGTHS	WEAKNESSES	OPPORTUNITIES	THREATS
e-Infrastructure and data operators		<ul style="list-style-type: none"> - Many e-Infrastructures are already implementing PIDs for data and contributors; - Some third-party services already exist, fostering further adoption 	<ul style="list-style-type: none"> - Tendency to use home-grown solutions; - Lack of interoperability between PIDs hinders adoption and reduces benefits to users 	<ul style="list-style-type: none"> - Build bridges between community-grown PID infrastructures; - Allow data exchange/access with interoperable frameworks 	Services that are not part of a global or cross-disciplinary layer became silos and threaten wide adoption of PIDs (and third-party services)
Scientific and research communities		Researchers are motivated by discoverability—of their own work and that of others that helps support their research. They embrace PIDs solutions that support discoverability and facilitates their work to be cited and re-used	The lack of an interoperable approach to attribution, citation, and measurement of re-use of data reduces the incentives for researchers to engage with PID initiatives	Best practices for PID systems supporting attribution and access exist in many disciplines. These examples can be leveraged to build a globally interoperable PID infrastructure	Disciplinary silo solutions, if easy to use and well established for a given community, present a barrier for adoption of interoperable systems and global solutions
Scholarly communication community	Libraries	Experience with linking diverse source of information, knowledge databases and standards	Most services often focused on text publications, missing link to datasets and other scholarly materials	Emerging intention to collaborate with other stakeholders to added value to each other services	Limited access to global PID infrastructures and limited resources to contribute to it may lead to marginalization and loss of content and valuable experience.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	32/57

		STRENGTHS	WEAKNESSES	OPPORTUNITIES	THREATS
	Publishers	Early adopters and awareness of PID for scholarly materials (CrossRef) and authors (ORCID)	Varying level of awareness or interest in interoperability with data and technological barriers for its integration in scholarly communication	Capture PID for authors and non-textual material with journal articles, build links, through interoperable infrastructures, to other services.	Limited capture of data PID or lack of integration with PID systems may lead to "undiscoverable"/"dead-end" material undermining incentives for other stakeholders
Funders and Policy Makers		Fast-growing awareness and strong commitment to PID infrastructures underpinning data-intensive science	-Interoperable PID infrastructures not yet in the focus of policies/ funding - Ad-hoc national solutions sometimes favoured instead of global ones	- Interoperable and global PIDs underpins effective management of science - Incentives from PIDs result into more open science and return on investment.	Lack of co-ordinated support to global, participative, PID infrastructure and thus, opting for local non-interoperable solutions

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

7. GAPS

The SWOT analysis underlines crucial gaps in the current PID landscape. These gaps are described in this section and are one of the key results for the participative and consultative ODIN process. The interviews highlighted that timing is crucial to address these gaps now given that PID solutions are emerging but many lack interoperability. Section 8 will address some of the emerging initiatives that are already addressing these gaps and will be part of the ODIN later proposal for action.

Gap 1: There is only limited access to PID e-Infrastructures for small organisations. A large number of mostly small organisations are not yet participating to the emerging global international and interdisciplinary PID e-Infrastructure. PID systems need to be broadly implemented in order to have a great impact on discoverability and open science. Researchers at those organisations (and their outputs) are “invisible” in the new paradigm. Access to e-Infrastructures and PID systems has to be facilitated across large and small organizations, and across borders.

Gap 2: Some research communities have little to no experience with interoperable PIDs for data and contributors. For a large number of research communities, experience with PIDs has been largely with DOIs for publications. PIDs for authors or non-textual scholarly materials, and in particular data, are neither established nor emerging.

Gap 3: Local, tailored, PID systems, with no interoperable options, are emerging. While local solutions can be customized to the needs of a specific community, they are often designed without interoperability in mind. They cannot link with other systems and therefore cannot overcome disciplinary or national boundaries. This is a

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

severe barrier to discoverability and access and thus widespread adoption of PID infrastructures.

Gap 4: There is a lack of support and funding to implement international interoperable PID solutions. Organizations aspiring to participate to a global PID e-Infrastructure may lack resources to implement PID systems and/or ways to link their local services with wider interoperable systems, exposing their data. There has to be dedicated support to encourage institutions to participate in global PID e-Infrastructures.

Gap 5: Methods and tools to track re-use of research data and other scholarly materials are lacking. Currently researchers cannot track re-use of the data they share. PID systems can enable tools that track the re-use of these datasets and further incentivize data sharing through attribution and citation.

Gap 6: Policies to encourage data sharing and acknowledge data re-use in research assessment are not yet widespread. A major component in the grant-making process is an assessment of the impact of prior work of a researcher, an institution or a program. Publishing and sharing data, which will underpin e-Science, is not yet included in this assessment, as there are no metrics to measure the use and reuse of datasets (Gap 5).

Gap 7: Reliable discovery services for research data and non-text based scholarly materials are missing. That little research data openly available today is currently accessed mostly via disciplinary data repositories or ad-hoc interfaces. This can hinder cross-disciplinary discovery and re-use of research data. Integration of interoperable PID solutions in existing systems can enable cross-platform discovery services and third-party search portals.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Gap 8: Incentives for making datasets re-usable are unclear or missing. Data repositories offer citation recommendations for research datasets, however, research data are still rarely cited. Data citation needs to be facilitated and incentive systems for data citation need to be agreed on and developed, to trigger a virtuous circle of data sharing.

Gap 9: Value-added services that can incentivize citation and open science cannot be built for lack of a widespread, interoperable, PID infrastructure. Although all actors are highly aware that value-added services for the scholarly community are needed, the lack of widely adopted interoperable PIDs makes it difficult if not impossible to create local or global (third-party) value-added services.

Gap 10: Unique attribution and linking between researchers, their scholarly materials and funding, is just not possible without a collaborative adoption of global and interoperable PID systems. The real social fabric of science depends on citation of scholarly materials, attribution to contributors, and eventual link with the grants that funded the work. There is to date no homogeneous and widespread adopted solution. Still, DataCite, ORCID and FundRef are emerging examples of participative community solutions to those systemic issues.

8. FROM THE GAP ANALYSIS TO A DRAFT ROADMAP

The interviews conducted by ODIN showed that PID management is a pressing concern for several of the stakeholder group and increasingly being implemented, across disciplines and organisations. This is not only about minting PIDs at a particular point in the research process, but also about establishing a trustworthy

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	36/57

system for managing, and exposing, PIDs, with the relevant interconnections and understanding of roles, responsibilities and costs.

In one example, third-parties are creating services to track re-use of data (Gap 5). First initiatives, such as Impact Story, integrate not only citation information, but also usage statistics (e.g., downloads). The Data Citation Index¹⁷ is under development (for release in 2014), offering an opportunity better data citation tracking based on DOI access statistics. More work and testing are needed to explore the methods and utility of these statistics, and how other stakeholders can contribute, but progress is clear. In common between these tools is the fundamental need for a reliable PID infrastructure. These services require an integration of infrastructure and services across disciplines. Timing is crucial for development of standards for an interoperable PID framework to support these tools, which in turn support sharing and open science.

Regarding adoption of international PID systems (Gap 3), the biomedical community with its large production of research papers and materials has long been one of the forerunners. The community uses PubMed identifiers to point to their publications or Protein Data Bank or GenBank accession numbers to their datasets, just to name a few examples. Infrastructure providers including Europe PubMedCentral now build on these pioneering infrastructures and have implemented ORCID as their author identifier¹⁸. Several more organizations are poised to follow suit.

Open Science is an emerging practice – researchers are sharing more and more materials. Policy makers demand more open science practices. Content needs to be

¹⁷ http://wokinfo.com/products_tools/multidisciplinary/dci/

¹⁸ <http://blog.europepmc.org/2013/08/linking-articles-available-in-europe.html>

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

linked and enriched so that materials can be found and reused: silos need to be combined, linked or integrated to avoid further isolation. To unlock the potential of the research materials, sharing needs to happen according to PID and data exchange standards. The Research Data Alliance (RDA)¹⁹, a global effort to enable and support research data sharing across stakeholder groups, has started working and interest groups in that regard, as part of the wider convening power of all stakeholders in a unique bottom-up approach.

To support Open Science, data sharing must be an integral/embedded part of the research process. As mentioned before, this means that services need to do a better job - allowing seamless access from paper to data (while submitting OR retrieving it). This would improve discoverability, reducing burden on researchers—while also demonstrating the value of open science to the research process. Value added services are required to enhance the search and analysis processes. This is not only of benefit for the users of the data, but feeds back to the data producers who see and track data use and shows that their extra effort to share data was worth the effort.

The research conducted by ODIN underlined that there is a path to this vision. Collaborations across stakeholder groups are needed to build bridges and to enhance functionalities, as recommended in the DIGOIDUNA report. Our recommended actions are not simply to develop new working groups or task forces, but to join forces to build integrated services. This could for example follow successful examples, such as CrossRef, which built on a multi-stakeholder approach to serve the global scholarly communication community addressing a concrete issue. DataCite takes a similar approach and developed a membership model with e-Infrastructure providers and libraries. ORCID builds on a multi-stakeholder approach,

¹⁹ <https://rd-alliance.org/node>

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

with stakeholders representing researchers, funding organizations, professional associations, universities and other research organizations, publishers, and third-party vendors, all of whom are critically involved in the success of an international PID for researchers and contributors.

They all provide an example of how organizations can work together to address this gap and the potential benefit of more general adoption of interoperable PID systems. Those need to be encouraged to mature and pervasively scale.

9. RECOMMENDATIONS FOR FUTURE ACTIONS: DRAFT ROADMAP

A set of targeted, concrete, actions can address each of the gaps identified leading to the implementation of an interoperable PID layer that supports open science. This will require action by all stakeholders. The following concise actions have been identified through an iterative process including external input during the first year (phase 2 based on desk research and the kick-off event) and been revised after the interviewees comments (phase 4). They frame a starting point for the second year of ODIN during which those will be further explored and validated (see Table 8 for an overview of gaps and actions).

Gap 1: There is only limited access to PID e-Infrastructures for small organisations.

→ **Action 1:** *Lower access barriers for institutions to participate to interoperable global PID e-Infrastructures, through appropriate agreements between institutions, fostering collaborations and with the support of national/international bodies.*

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Gap 2: Some research communities have little to no experience with interoperable PIDs for data and contributors.

→ ***Action 2:** Support those scientific communities without existing PID solutions to participate to existing interoperable PID frameworks, while tailoring interfaces to the specificity of the community.*

Gap 3: Local, tailored, PID systems, with no interoperable options, are emerging.

→ ***Action 3:** Facilitate interoperability between stakeholders with community-specific, institution-specific or national PIDs solutions and emerging global open solutions.*

Gap 4: There is a lack of support and funding to implement international interoperable PID solutions.

→ ***Action 4:** Provide (seed) funding to ease local participation and access to emerging PID infrastructures.*

Gap 5: Methods and tools to track re-use of research data and other scholarly materials are lacking.

→ ***Action 5:** Develop an interoperable PID infrastructure that supports development of third-party tools for discoverability, impact assessment, and other value added services.*

Gap 6: Policies to encourage data sharing and acknowledge data re-use in research assessment are not yet widespread.

→ ***Action 6:** Design policies to elevate data to a key indicator in research assessment, with appropriate attribution to their creators and curators, through implementation and usage of open and interoperable PIDs.*

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Gap 7: Reliable discovery services for research data and non-text based scholarly materials are missing.

→ **Action 7:** *Harmonize formats and APIs, so that information from emerging and existing PID frameworks can be exposed and mutually enriched, while enabling third-party discovery services.*

Gap 8: Incentives for making datasets re-usable are unclear or missing.

→ **Action 8:** *Design appropriate incentive systems to pervade research evaluation, e.g. citation mechanisms based on PIDs for data, linked to PIDs for contributors.*

Gap 9: Value-added services that can incentivize citation and open science cannot be built for lack of a widespread, interoperable, PID infrastructure.

→ **Action 9:** *Assure that a trusted, open and sustainable interoperable PID infrastructure is established with ease of participation of third-parties.*

Gap 10: Unique attribution and linking between researchers, their scholarly materials and funding is just not possible, without a collaborative adoption of global and interoperable PID systems.

→ **Action 10:** *Establish a participative framework with PIDs for contributors and materials, where any participant can expose information, enriching the entire e-Infrastructure.*

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	41/57

Table 8: ODIN overview of gaps and corresponding actions focused on actors and beneficiaries

GAP	ACTION	ACTOR	BENEFICIARY
Gap 1: There is only limited access to PID e-Infrastructures for small organisations.	Action 1: Lower access barriers for institutions to participate to interoperable global PID e-Infrastructures, through appropriate agreements between institutions, fostering collaborations and with the support of national/international bodies.	e-Infrastructure providers; funders/ policy makers	e-Infrastructure providers long tail of research institutions
Gap 2: Some research communities have little to no experience with interoperable PIDs for data and contributors.	Action 2: Support those scientific communities without existing PID solutions to participate to existing interoperable PID frameworks, while tailoring interfaces to the specificity of the community.	e-Infrastructure providers; funders/ policy makers	research communities

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	42/57

GAP	ACTION	ACTOR	BENEFICIARY
Gap 3: Local, tailored, PID systems, with no interoperable options, are emerging.	Action 3: Facilitate interoperability between stakeholders with community-specific, institution-specific or national PIDs solutions and emerging global open solutions.	librarians; e-Infrastructure providers	e-Infrastructure providers, researchers, funders/ policy makers
Gap 4: There is a lack of support and funding to implement international interoperable PID solutions.	Action 4: Provide (seed) funding to ease local participation and access to emerging PID infrastructures.	funders/policy makers	e-Infrastructure providers
Gap 5: Methods and tools to track re-use of research data and other scholarly materials are lacking.	Action 5: Develop an interoperable PID infrastructure that supports development of third-party tools for discoverability, impact assessment, and other value added services.	e-Infrastructure providers (in particular 3rd party services for impact assessment)	research communities, funders/policy makers,
Gap 6: Policies to encourage data sharing and acknowledge data re-use in research assessment are not yet widespread.	Action 6: Design policies to elevate data to a key indicator in research assessment, with appropriate attribution to their creators and curators, through implementation and usage of open and interoperable PIDs.	funders/ policy makers, e-Infrastructure providers, librarians	researchers

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	43/57

GAP	ACTION	ACTOR	BENEFICIARY
Gap 7: Reliable discovery services for research data and non-text based scholarly materials are missing.	Action 7: Harmonize formats and APIs, so that information from emerging and existing PID frameworks can be exposed and mutually enriched, while enabling third-party discovery services.	librarians, e-Infrastructure providers	researchers, funders/ policy makers
Gap 8: Incentives for making datasets re-usable are unclear or missing.	Action 8: Design appropriate incentive systems to pervade research evaluation, e.g. citation mechanisms based on PIDs for data, linked to PIDs for contributors.	librarians, publishers, funders/ policy makers	researchers, funders/ policy makers
Gap 9: Value-added services that can incentivize citation and open science cannot be built for lack of a widespread, interoperable, PID infrastructure.	Action 9: Assure that a trusted, open and sustainable interoperable PID infrastructure is established with ease of participation of third-parties.	funders/ policy makers, e-Infrastructure providers (especially 3rd party services)	e-Infrastructure providers

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy	Dissemination level: PU	
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)	Version: 1_0 Final	44/57

GAP	ACTION	ACTOR	BENEFICIARY
Gap 10: Unique attribution and linking between researchers, their scholarly materials and funding is just not possible, without a collaborative adoption of global and interoperable PID systems.	Action 10: Establish a participative framework with PIDs for contributors and materials, where any participant can expose information, enriching the entire e-Infrastructure.	e-infrastructure providers, librarians, funders/ policy makers	researchers, funders/ policy makers

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

All stakeholder groups must be engaged in the process of implementing the missing thin layer for a global PID e-Infrastructure, from researchers to funders and policy makers. This is a challenge at a European and global level.

Our analysis indicates the need to engage existing early adopters of PID infrastructure, across all stakeholder groups, in the process of organizing and implementing an interoperable, open and participative framework. This includes IT systems, data centres, libraries, and publishers, universities and research centres, professional associations, funders, and third-party vendors—and researchers themselves. Every party has responsibilities to ensure the development and longevity of this framework. The analysis also shows that all parties are beneficiaries of the process, making engagement and participation a win-win situation.

10. CONSIDERATIONS FOR THE FUTURE ACTIONS AND ODIN 2ND YEAR

These findings presented here allow charting a clear course for the second and final year of the ODIN project. ODIN will engage more closely with early adopters across the major stakeholders to prepare for these actions, starting with our 1st year event and the hands-on CodeFest, where real interoperable products might be conceived and tested on the existing open APIs of DataCite, ORCID and beyond. The key process will be further consultation and refining of the actions, as well as sounding each stakeholder group on their acceptability, in the longer view of building a wide consensus. To this extent we detail a first picture of the main responsibilities of each stakeholder and propose how future collaborations could develop.

- **e-Infrastructure providers and data centres:** *build on community infrastructures and enhance them with standardized approaches.*

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

It is important to build on existing solutions. Local identifier systems shall be further developed and shifted to an interdisciplinary and interoperable PID infrastructure. There are already some best practice examples of data centres using PIDs and collaborating with publishers to provide improved discoverability for datasets by direct linkage to an article on a publisher platform. Cooperation with third-party services could furthermore facilitate citation (and re-use tracking) of deposited material. Data centres are considered a trustworthy place to preserve data and thus are an important partner for other stakeholders who can provide services on top of the data centre metadata.

- **Researchers:** *best practice examples exist in many disciplines - important to learn from these and transfer these to the own community*

Early-adopter scientific communities have developed locally interoperable PID systems in their workflow for citing papers and sharing and citing research data. Communities that are less advanced in the management of PIDs can learn from these best practices. Furthermore, this can help build more bridges between disciplines.

- **Scholarly communication communities**

- **Libraries:** *Collaborate with other stakeholders and offer added value to their services and expose their materials*

PID systems offer opportunities to provide more comprehensive information to users. Information can be linked via PID systems across domains (authors, papers, datasets, affiliations, service activities), offering more possibilities for information retrieval and a higher chance for researchers to find useful material and increase the incentives for re-use of open science materials.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

- **Publishers: *Leverage existing interactions with researchers to accelerate adoption of global PID infrastructures.***

Based on the existing interaction with research communities, publishers are in a good position to engage with research communities. Most of them traditionally assign interoperable PIDs (DOIs) to their own materials. They should grow this practice to foster participative adoption and contribution of PIDs for authors and non-textual scholarly material. This tradition of engagement should expand to collaborations with other stakeholders.

- **Funders, Policy makers: *Support and require interoperable and global PIDs infrastructures to leverage for impact assessment and data sharing across national and disciplinary boundaries.***

Funders expect the outputs of funded research to be openly available. PIDs help to make these results discoverable and citable and in turn support program evaluation. Such analysis enables funders to make appropriate decisions, PID systems integrated into funding workflows can support researchers in reporting their work.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

11. SUMMARY

This document has presented a stakeholder-specific SWOT analysis of PID e-Infrastructures and enumerated 10 serious gaps where the lack of a PIDs global solution is threatening the very fabric of data-intensive open e-Science. We have suggested some targeted actions, to address each of those gaps, clearly identifying the roles of each of the key stakeholders in connected communities: e-Infrastructures, research, scholarly communication and funding agencies. As an overarching summary, we can abstract three main tiers, underpinning the roadmap to address the overarching gaps identified in our analysis:

1. Deliver an interoperable PID layer for data and contributors of the highest quality, fostering participation of all existing early adopters. A high standard is crucial to enable adoption and assure development of third-party value-added services; these, in turn, support incentives for researchers and others parties to adopt PIDs.
2. Promote and support multi-stakeholder research on missing aspects of a global PID e-Infrastructure. Dialogue between stakeholders identifies features in need of development by a party, benefiting the other, furthering third-party adoptions. Examples are advanced disambiguation systems, PIDs for entities beyond researchers, global integration of PIDs to a longer tail of scholarly.
3. Design and implement sustainable and participative business models to lower the participation barrier for entities to embrace a global PID e-Infrastructure, while assuring a resilient operation and prevailing openness.

The second year of ODIN will elaborate on those three overarching targets, through further consultation and consensus-building, around very concrete actions. We aim to concentrate on the potential acceptance from each stakeholder community of the actions and the overall targets, and deliver a final roadmap for implementation, possibly in the Horizon 2020 framework and timescale.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

12. APPENDIX 1: INTERVIEWS

ODIN conducted interviews with experts in the field of PID management and digital scholarly communication in May 2013. The list of potential interviewees was shared within the ODIN project partners to determine the best possible coverage of all stakeholders and disciplines. It was expected that the situation and challenges described were to vary so that an extensive coverage was targeted. Our background research findings were sent to the interviewees in advance. The concise 2 page briefing paper is included below (Section 12.1). In addition, a semi-structured questionnaire was developed for the interviews (Section 12.2). Note that the questionnaire was not applied strictly - it was made sure that all the questions were addressed during the course of the conversation, but not necessarily in the same order.

12.1. Sheet for the interviewees to comment on

What this is about?

There will be no Open Science without researchers' buy-in. Recent studies highlight that researchers' engagement is still rather limited or uneven across disciplines and subfields and that incentives need to reach a tipping point in open sharing and reuse of scholarly materials. The most effective incentives relate to scholarly metrics: those used in career and research assessments, as well as new, emerging, informal ones.

The ODIN project is based on the vision that the actual implementation and success of such incentive mechanisms is depending on a reliable infrastructure and service framework with persistent identification. Today, it appears that there are several gaps to realize this vision. This can be seen in the first ODIN results, based on desk research and the kick-off event in the first year. These first steps of the project

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

resulted in concise hypotheses and a draft roadmap (enclosed in the end of this document).

ODIN wants to dig deeper and now conducts a more detailed gap analysis to detect current obstacles and missing actions which are required for an interoperable persistent identification framework.

As part of this, ODIN conducts interviews with a series of experts active in scholarly communication, to detect the most prominent gaps for the individual stakeholder groups and their impact. Furthermore, we seek expert's input on the first results and the draft roadmap. The multi-stakeholder approach shall ensure the broad, but detailed applicability of the outcome.

This document provides you, the interviewee, with a short background about the work of the ODIN project and its approach. It shall help you and us setting the scenes for our conversation about persistent identification and the facilitation of Open Science. This will then feed into and sharpen ODIN's vision and report to the EC.

The ODIN Project

ODIN is a 2-year funded project by the European Commission under the umbrella of the 7th Framework Programme. ODIN's partners are ANDS (Australian National Data Service), arXiv, British Library, CERN, DataCite, Dryad and ORCID. ODIN focuses on the interoperability and future of persistent identifiers for data and researchers across the entire ecosystem of scholarly communication. ODIN partners represent the broad spectrum of stakeholders active in scholarly communication, with no disciplinary boundaries, with partners of the exact sciences, the natural sciences and the social sciences and humanities.

www.odin-project.eu

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

The interviews

The interviews will inquire about your past, current and forthcoming work with persistent identification for scholarly materials and their authors/producers/contributors. We would like to understand the obstacles you encountered or expect to come and what is needed to overcome them. Of particular interest is the persistent link between authors, publications and research data. Depending on your background (e.g. publisher or funding body), your experience might be very different.

In addition to this first part, ODIN seeks your input on our first results. We have already gathered a series of hypotheses and a draft roadmap which are enclosed in this document (in a very concise form, as brief hypotheses and a draft roadmap). At the end of the interview, we would like you, as the experts in the field, to comment on these and, if applicable, highlight the impact on your individual work.

PART II:

ODIN hypotheses

Based on our kick-off event and desk research we have phrased some hypothesis of the main obstacles and gaps that exist. Also, a concise roadmap has been drafted. We are interested in your feedback on these.

What did we miss, which one is particularly important for your work and why?

What actions are needed to solve the issue in the particular work environment?

What is your opinion on the draft roadmap?

- In some disciplines, researchers are not aware how and where they could share their data.

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Action: More dedicated training mechanisms, take advantage of the visibility opportunities and address the needs of sensitive data that might be shared.

- In particular smaller institutions or infrastructure providers are not aware/lack the expertise and funding to implement international standards or value-added services.

Action: Better training and collaborations are needed to take them onboard.

- Researchers and funders would like to know how often scholarly materials, such as datasets are being cited or reused. The impact of the individual materials needs to be assessed. But tracking of such resources is not widespread. The first solutions are emerging, but adaption is challenging.

Action: Interoperable systems are needed to allow tracking.

- Researchers would like to get credit for sharing their scholarly materials openly. Research assessment and reward systems are focused on scholarly articles, not open science products.

Action: Recognize such materials, as it is already done in the example of the NSF Grant Proposal Guidelines.

- Publishers would like to support a scholarly communication market in which open data sharing counts. It is difficult to find trustworthy, sustainable data repositories to collaborate with. Such collaborations are crucial to establish the link between author-articles-data.

Action: Establish trustworthy repositories and persistent links between author and materials

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

- Researchers would like to have better access to their and their colleagues research materials for future reference. It lacks workflows and portals to gather and display information about such scholarly materials across disciplines. This is mainly related to interoperability issues across platforms.

Action: Libraries, repository providers and publishers need to collaborate and facilitate seamless access

- To facilitate reuse such scholarly materials should be a citable object. Currently, the adoption of DOIs as a citable persistent identifier is emerging in many data centres, but data citation practices are not established in many disciplines.

Action: Improve the support and training in that regard.

Roadmap:

ODIN suggests three main tiers (high quality services, global multi-stakeholder approach, sustainable business models) to support a better integration of persistent identification in scholarly communication, in particular to support an integration of different identifier schemes across stakeholder groups.

1. The delivery of high-quality services is crucial. Such could be used by third parties to build additional - value adding- services. This will feed back to the overall PID usage and will enlarge the user basis of the individual persistent identifiers contributing to a global integrative and interoperable system.

2. It requires a European and global multi-stakeholder approach to tackle the challenges of de-duplication and disambiguation. These are, in particular, relevant to

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

ensure that metrics based on such implementations can be accurate and used for future research assessments.

3. It is needed to study the implementations on the long term, aiming for sustainable and participative business model

12.2. Questionnaire

About the interviewee:

Institution/Company:

Position:

Responsibility/Job description:

Specific to a scientific discipline, if applicable:

Question 1:

Please explain where and how you deal with PID in your work?

Question 2:

Do you use PIDs for author? If yes – which ones?

Do you use PIDs for scholarly materials? [DOI, Handle, ARK...]

Question 3:

How far/advanced is the implementation?

What about the triangle - author/publication/data... , are you aware of actions or participate in establishing/preserving the connections?

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Question 4:

What are the main challenges for you when dealing with PIDs for authors and scholarly materials? What is the strongest barrier?

- Infrastructural/technical/IT?
- Organisational? For example: to find interesting collaborators?
- Budgetary?

Question 5:

Where would you need better support now? [or in 5years time]

Question 6:

What would be your ideal case scenario for PID implementation within 5 years [10years]? How would this impact/have impacted the work of your institution?

If applicable:

- Does your platform/implementation facilitate reuse/citation of scholarly materials other than text/papers, i.e. research data?
- Do you intent do show citation stats for other materials, i.e. research data?
- Do you intent to consider usage stats as well? If yes, what exactly do you have in mind?
- If you work predominantly for a single discipline – is there a plan to expand to other disciplines? If yes, how/where... what are the main barriers?

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

Final questions on ODIN results.

Do you agree with the overall ODIN results summary?

Any remarks/additions/corrections?

Concerning the scenarios – does any of them apply to you particularly?

What are your thoughts on the roadmap?

	D5.1 Gap analysis and draft roadmap		
	WP5: Strategy		Dissemination level: PU
	Authors: Sünje Dallmeier-Tiessen (CERN), Patricia Herterich (CERN), Salvatore Mele (CERN)		Version: 1_0 Final

13. APPENDIX 2: LIST OF ABBREVIATIONS

EC	European Commission
HEP	High Energy Physics
HSS	Humanities and Social Sciences
ISNI	International Standard Name Identifier
NSF	National Science Foundation
ODE	Opportunities for Data Exchange
PID	Persistent Identifier
RDA	Research Data Alliance
URN	Uniform Resource Name
WP	Work Package