DIGOIDUNA

Digital Object Identifiers and Unique Authors Identifiers to enable services for data quality assessment, provenance and access

REPORT FOR EXPERTS FEEDBACK

UNIVERSITÀ DEGLI STUDI DI TRENTO

SMART 2010/0054

_____

Trento, October 2011

Table of Contents

Executive Summary

The strategic role of e-Science infrastructures as a crucial asset enabling European scientific advances and innovation has become a central issue in the digital agenda for Europe. These infrastructures provide a distributed virtual environment where scientific resources can be accessed, shared and exploited supporting global research collaboration and opening new frontiers to data processing and scientific discovery.

In this context, the functionality to unambiguously locate and access digital objects, associate them with the related authors and other relevant entities in the scientific production chain (e.g. collaborators, institutions, projects) becomes essential. Digital identifiers play a fundamental role in this respect, underpinning added-value services such as data quality assessment, provenance and access.

The objective of the DIGOIDUNA study is to support policy makers at European, Member State and research institution levels in assessing impacts and understanding the opportunities and challenges connected to managing *digital identifiers* within the context of *scientific data e-infrastructures* (SDIs), providing instruments to support decision-making on solutions that will have a long-lasting influence on scientific research and on the long term access, preservation and integration of valuable data and knowledge assets held within the sector.

This document reports the results of the DIGOIDUNA study. The report is focused on three key objectives: 1) analyzing the fundamental role of identifiers as enablers of value in e-infrastructures and presenting forward looking scenarios as examples of the benefits of a systematic usage of identifiers for digital objects and authors to locate and integrate information from multiple sources; 2) reporting the results of the analysis of the Strengths, Weaknesses, Opportunities and Threats (SWOT) associated with establishing in Europe an open, dynamic and sustainable governance of e-infrastructure using identifiers for digital objects and authors; 3) presenting the main challenges and recommendations which European Commission and other relevant stakeholders should address to develop an open and sustainable e-infrastructure for locators of digital objects and identifiers of authors supporting scientific information access, curation and preservation.

The general thesis, confirmed by the study, is that at a technical and informational level identifiers and their management within each localised activity constitute the threads from which the very fabric of an emergent digital global knowledge infrastructure is woven. The need for awareness and coordination from the outset is vital. This is especially true in the design of e-infrastructures ensuring the management of data and information created and shared across national, organizational, disciplinary, cultural and technological frontiers, as this requires to move from local centralized authority-based solutions to cross-boundary interconnected solutions, raising a whole set of new issues and challenges which deal with economic, societal, policy and implementation issues.

The overall conclusion of the study is that actions can and should be taken to enhance the value that identifiers infrastructures can bring to extract the maximum value from the overall industry of scientific research. In the widest sense, this involves connecting those identifiers used within stakeholders: i.e., through SDIs, knowledge transfer through established routes evolved from traditional scholarly publishing, connecting across the web as a whole, as well as with government and funding bodies. This is a situation that is already being currently addressed in various ways, but often with different starting points and goals. With proactive intervention and efficient coordination amongst key stakeholders, an identifier

infrastructure offers tremendous potential for accruing and harvesting past investments into European R&D, and so maximising delivery on the value captured within its SDIs.

Recommendations are organised along two distinct trajectories.

- The first trajectory is in a way that takes account of the lifecycle of the elements in an ecosystem that addresses the issues within the scope of the study. The <u>first</u> overall recommendation is to **define a common agenda** for further analysis, design and implementation of initiatives and monitoring of progress by means of an overall conceptual model and framework for action. The <u>second</u> is to address the need to **bootstrap an identifier eco-system** that is firmly focussed on delivering near-term demonstrable value from SDIs through engagement with important *existing* initiatives. The <u>third</u> is to concurrently promote awareness and proactivity levels to **mobilise Europe's extensive resources and skills** within research arena to effect change, and the <u>fourth</u> directly addresses the **delivery of ongoing sustainable value** from SDIs, through a combination of measures aimed at creating models and leveraging engagements to secure ongoing innovation, coordination and interoperability.
- The second trajectory is organised in terms of measures aimed at addressing specific drivers, *viz.* creating a critical mass of coordinated actions, securing long-term viability, creating demand and the conditions for exploitation, and guidance for technical standardisation and RTD-related activity.

The report is organised in three main sections, which correspond to the three objectives described above. Background material is made available on the DIGOIDUNA website (http://www.digoiduna.eu/).

The current version of the report might be slightly adapted to include the input given in the DIGOIDUNA final workshop planned on December 2nd, 2011 at the British Library Conference Centre, 96 Euston Road, London (UK).

1 Introduction

The European knowledge society is entering a new phase of development where ICT is providing the key basic infrastructures for all vital social and economic processes and is the most influential key technology in most innovations across all industries. All private and public services are being provided through and shaped by these infrastructures[1]. On the other hand, innovation closely depends on rapid scientific and technological advancement. Scientific research has become extremely data intensive and it is getting more and more interdisciplinary and internationally performed. Therefore, *the use of high-capacity computing and research infrastructures are increasingly playing a crucial role in fostering scientific progress advances*, by promoting the development of efficient research and innovation environments. It follows that the progress in ICT and the development of new methods for exploiting digital resources is indispensable to address key challenges within what is termed in this report *research production*: the outputs of the business of conducting scientific research.

Research production plays a defining role in our economy, which in turn means that SDIs provide a critical infrastructure for the global economy. *Just as physical networks connect physical machines, digital identifiers (DIs) connect the digital data and content held within them*. Machines and their connectivity are essential for the productivity of e-Science, but so also is the connectivity between people on these networks, creating another dimension of ICT users to reach through DIs: i.e. researchers and practitioners, funding agencies and wider portions of the societal tissue. But this is only part of the story. DIs have the potential to connect contents to authors and authors to their institutions, research projects and potentially any other relevant entity in the production value chain of scholarly content. In this respect, DIs could tighten the threads of an articulated network of connections within the represented research ecosystem, opening new prospects for advanced information integration services in e-Science. The successful implementation and sustainability of DIs within SDIs thus becomes an issue of central importance.

The objective of the DIGOIDUNA study is to support policy makers at European and Member State level to understand the opportunities and challenges connected to managing DIs within the context of establishing SDIs, and provide instruments to support decision-making on solutions that will have a long-lasting impact on scientific research and on the long term access, preservation and integration of valuable data and knowledge assets held within the sector.

In the Interim Report of the study, the basic meaning and functioning of DIs have been explained, the key terms introduced and defined, the complex ecosystem in which DIs are used analysed and the current state of affairs depicted, mostly relying on the available literature and on interviews with experts in the domain[2].

Based on this material, this Report focuses on a summary-level assessment of the current situation, on the identification of future-looking scenarios which can be enabled through a suitable framework for managing DIs for e-Infrastructures, and on the diagnosis of the gap which must be filled to move from the current to the ideal situation. The final part contains a structured list of recommendations, which are proposed to fill the identified gap. This list is organized around four key actions: (1) **defining** a framework for further analysis, actions and monitoring of progress by means of an overall conceptual model and framework for

---

[1] See "Orientations for EU ICT R&D & Innovation beyond 2013 - 10 Key Recommendations - Objectives, rationale and content" (http://cordis.europa.eu/fp7/ict/istag/documents/istag_key_recommendations_beyond_2013_full.pdf)

[2] Part of the material presented in the Interim Report has been made available on the DIGOIDUNA website at http://www.digoiduna.eu/

action, that brings in the key yet diverse elements pertaining to SDIs; (2) **bootstrapping** an identifier eco-system that is firmly focussed on delivering near-term demonstrable value from SDIs through engagement with important existing initiatives; (3) **mobilising** Europe's extensive skills base within research arena to effect change, (4) **delivering** an ongoing **sustainable value** from SDIs, through a combination of measures designed to create models leverage engagements to ongoing innovation, coordination and interoperability.

2   Digital Identifiers: Empowering e-Science

Digital Identifiers and their role

The use of identifiers has always played a fundamental role in information management, even before the advent of computer-based information systems. We just need to think of identifiers for books in traditional libraries, for people on documents, for buildings in cities. From their early uses, identifiers were always employed both for distinguishing similar objects (for examples, two or more people with the same name) and in some cases for making access to the object simpler and more efficient (for example, a book tag in a library or a street address for a building). The most fundamental property that makes an identifier useful is its unique *binding* to its referent. Therefore, from the beginning, the main success factor of an identifier systems is fundamentally based around the successful implementation of *governance procedures* of each authority controlling the levels of guarantee provided over accuracy, authenticity and reliability of the relationship between a given identifier (numbers or letters) and the entity it is intended to refer to, namely its *referent*.

The digitization of data and content has definitely increased the value provided by using identifiers, as they have become the key to precisely and effectively accessing digital content stored in information repositories. The paradigmatic example is their use as primary keys in relational database management systems, where they are used to distinguish a dataset record among millions of other records and to retrieve the record through structured queries. These digital repositories have mostly inherited the governance models and the management processes of the past. Indeed, they are typically managed under a single *authority*, whether a centrally and institutionally administered database system, a digital archive, or a library, or at national level such as a passport office.

The next phase was started by the advent and rapid adoption of Internet-based methods. The key difference is that now information repositories can be connected to each other, and this means that data and other content related to the same entity (either digital or non-digital ones, like a person) stored in different repositories have the potential to be interlinked and integrated. In this fully distributed, decentralized and heterogeneous environment of data production and consumption, digital identifiers are expected to provide *global keys for information access*, and as such are a critical enabling factor. The most striking example of this is the World Wide Web, where a global addressing system (based on *Uniform Resource Identifiers*, or URIs) allows to point to the same digital resource (e.g. a PDF file) from any location on the Web.
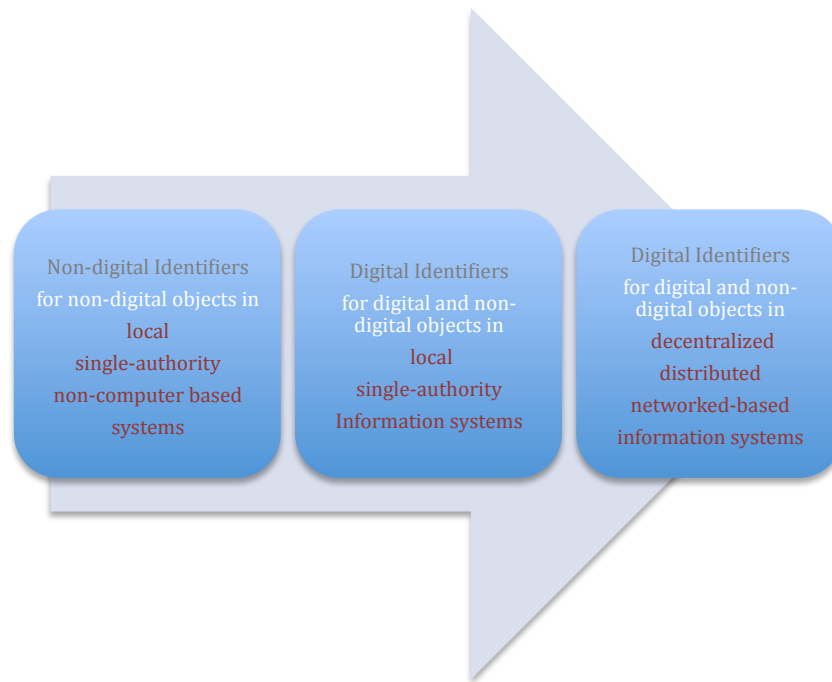
Figure 1: The evolution in the use of identifiers

If, on the one hand, this new scenario offers new exciting opportunities, on the other hand it raises new difficult challenges for identifier management systems. Unlike the shift from non-digital to digital information management systems, the shift from locally managed systems to a networked environment does not allow simple adaptation of the governance models and coordination practices which were developed in the past, as there is a need for a new kind of agreement on policies and procedures which cut across geographical, temporal, disciplinary, cultural, organizational and technological boundaries, without necessarily relying on a centralized system or authority. In other words, *what is needed today is a shift from a set of locally operated systems, each based on a single local authority, to a system where these authorities work in coordination to support the interchange and entire lifecycle of DIs*.

This general picture, which applies to most sectors of human activity, is particularly critical for scientific research. Indeed, research is by definition concerned with the creation, sharing and validation of experimental and sensor data, and this of course happens in a completely distributed and decentralized environment. For a scientist, being able to access and reuse good quality experimental data, integrating external data with her own, analyzing and correlating them, mining and distilling new findings is a substantial part of the research work. Accessibility, trustworthiness, quality and governance of data are essential needs, because they will ensure that virtual science communities can be formed, and that they can operate independently of geographic, political, institutional or temporal boundaries. In such a scenario, the consistent adoption and use of appropriate identifier schemes and systems is a necessary enabler.

This is the reason why assessing the impact that the usage of different schemes and systems for identifying digital objects and authors is an issue of the highest importance for policy makers, especially when one endorses a long term vision where large-scale ICT infrastructures for e-Science – in particular SDIs – will become the fundamental production environment for scientists. Such an assessment is very difficult, as it is requires to take into account not only the technical features of a scheme or system, but also a large number of complex socio-economical factors, including long-term sustainability, social acceptance, ownership, costs for users, privacy and so on. The remaining part of this section delimits the boundaries of

the DIGOIDUNA assessment, including some aspects of the methodology that has been adopted for the SWOT analysis used in the diagnosis of DIs deployment dynamics.

Identifiers: enablers of value in SDIs

Digital identifiers are universally recognized as situated at the root of extracting value from resources of information. Careful management and control of identifiers and maintenance of

> *"Digital Identifiers are the key to creating value from SDIs and their cost effective management."*

their relationship with the entities they refer to, whether digital entities within scientific data infrastructures, physical in the case of people, or conceptual entities in the case of precise meanings, constitutes the most efficient and accurate means of management of access to information. Since one of the main functions of scientific e-infrastructures is to support access, use and re-use of data, the role of DIs within SDIs appears crucial, as they represent the keys to access an increasingly volume of data and a mean to process, integrate, and correlate these data in a very efficient way.

In addition, DIs are an essential building block for enabling effective and efficient technical solutions towards the creation of value-added services on top of scientific data and contents. These services deal with many aspects of the e-Science landscape, including data and information access, knowledge discovery and citability, quality assessment and provenance, just to mention some of them. From this perspective, identifiers and metadata associated with them, allow to harness the potential benefits of the scientific data wave by increasing the interoperability of data infrastructures, facilitating the access to relevant, reliable and trustable information, increasing the trustworthiness of sources of information and related contents, revealing relevant links and dependences between data and solving reference ambiguities. To provide a very practical illustration of the enabling functions of digital object and author identifiers for SDIs, consider three examples of future e-Science scenarios envisioning their beneficial potential impact in different contexts of e-Science.

Scenario 1: Funding Agencies

Stefano is a Project Officer who works for the European Commission. He is filling the Impact Assessment Form, which is part of the final evaluation report of a project he has been assigned. The project adopted the new EC policy on persistent identifiers, which means that any relevant entity connected to the project has a unique identifier, which is compliant with the pan European e-Infrastructure for research data on climate change. This means that the project itself has an identifier, together with every contributor (scientists, PhD students, developers, experts, ...), every involved organization (universities, research institutes, companies, public bodies, ...). Any relevant entity is also geo-referenced. Stefano logs into the Project Impact Assessment portal and uses the project ID to collect all he needs to know to fill the form: the complete list of papers published or submitted for publication by the project team, a rich collection of metadata about them (type of publication, degree of inter-organizational collaboration, impact factors of journals, statistics on external and self-citations, and the like), a list of patents filed in the course of the project. He can browse the graph of EC funded projects and explore the history of past collaborations. Stefano can retrieve (and visualize on a map) other data, which prove the impact of the project in terms of jobs created within the participating partners, number and location of startup companies and compute an overall impact indicator.

The same happens for every EU funded projects, which in the mid terms allows policy makers to make much better informed decisions on the assignment of funds and provide a much more complete account to member states governments and European citizens.

Scenario 2: prepublication data sharing and citability

A research group at the Beijing Genomics Institute (BGI) has recently worked in collaboration with the University Medical Center Hamburg-Eppendorf researchers to understand a new virulent strain of E. Coli causing an outbreak of food poisoning that has killed 18 people In Europe. Using the genomic technology, scientists have determined the infectious strain and revealed the mechanisms of infection, facilitating the development of measures to control the spread of this epidemic. To maximise the utility of the findings to the research community and help scientists around Europe to find a treatment for the infection, the Chinese research group decides to release the genomic data into the public domain. The problem is twofold. First, the researchers need to find a way to claim their authorship on the dataset, protecting the intellectual property of the research. Secondly, they need to provide a mean to make the data citable before the publication of research papers on scientific journals that may take a significant amount of time. Therefore, the data of the genome's sequence of the bacterium are published with an identifier making the dataset a citable reference and providing an excellent example of a new form of data release that can provide benefits to different parties. On the one hand the researchers may benefit from creating a citable resource as it can later be used to reflect the impact of the dataset, improve the research collaboration and provide credit to authors before the publication of the results; on the other hand, the prepublication data sharing can maximize the collective benefit by speeding up the process of finding valuable solutions for threatening diseases.

Scenario 3: knowledge discovery and data integration

Ellen is a researcher at the Childhood Bilingualism Research Centre of Hong Kong and is conducting a psycholinguistic study on how children process words and sentences in Cantonese, English and Mandarin. During the review of the literature regarding this topic on a psychology database, she finds a paper published by a researcher of the University of Ottawa about how bilingual babies learn new words differently than monolingual babies. She finds the results reported in the paper very interesting for her study and decides to discover other works by the same author or other authors with similar research interests. She wants also to collect more information about the research activities and projects in which the author is (or has been) involved, the network of his collaborators and the institutions where he had previously worked in order to explore possibilities for research collaborations in future. A keyword-based research on a Web search engine shows that there are several authors and people with the same or similar name to that of the searched author and his name is recorded in several different ways. The author CV available on the Ottawa University Web site lists 3 main institutions where the author worked in the past and the list of publications seem incomplete and out of date. At a first glance, it seems to Ellen that finding exhaustive and integrated information about the author is not a trivial task. Fortunately, she remembers of a new knowledge discovery system for e-Science, called XYZ, based on unique author and object identifiers. Looking at the paper she is interested in, she notices that the paper reports such an identifier for the author and another identifier to

identify the paper itself. Using the resolution system offered by XYZ and the author ID, Ellen is able to disambiguate the name of the author and find all the relationships between the author, his publications and scientific contributions. In particular, she finds the complete lists of the papers published by the author and his co-authors, (including open access papers), the links to papers on the same research topic and links to the datasets that are made available by the research project funding the study and an integrated profile of the research activities, collaborations and citation metrics about the author. She also finds many references to track online contributions about the research topic, such as scientific blogging and community efforts.

The new challenges of cross-boundary solutions

The increased awareness that identifiers are a fundamental building block in the design of e-infrastructures has revealed another aspect of the use of identifiers, which deals with the globalization of e-Science. Science has become

> *"Moving from a local view to cross-boundary interconnected solutions poses new challenges for managing identifiers."*

increasingly based on open and cross-border collaboration between researchers and practitioners from any discipline across the world. This process has revealed the need of a proper scientific infrastructure to ensure permanent access, data sharing, provenance and quality assessment of data and information created and managed across national, organizational, disciplinary, cultural and technological boundaries. These issues call for moving beyond local centralized authority-based solutions to embrace cross-boundary interconnected solutions raising a whole set of new issues and challenges dealing with economic, societal, policy and implementation aspects in terms of the impact of using identifiers in complex e-Science eco-systems. Before analyzing in more details the implications of this paradigm shift, it is worth to mention some pioneering efforts, which have tried to provide an identifier solution to the cross-boundary challenges.
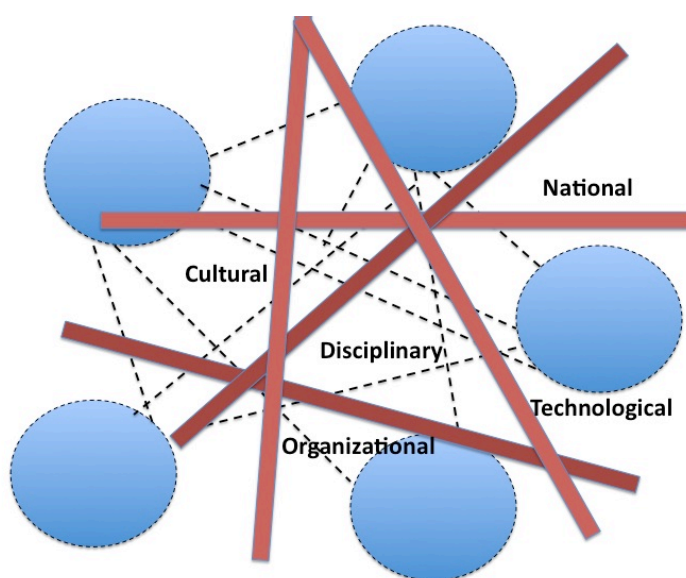

Figure 2: From local to cross-boundary solutions

The current state

Currently, two main types of approach aim at providing an integrating solution, which goes beyond the boundaries of a single institutional or commercial system and address the identification issue from a cross-system, cross-border, inter-organizational perspective. The most consolidated approach, termed here the **Persistent Identifiers (PID) approach**, is currently adopted and endorsed by the vast majority of significant stakeholders in the production value chain of scholarly content, including national libraries, publishers, research institutions and data centres. The strongest motivation behind this paradigm is the requirement of enabling trusted naming authorities, ensuring long term preservation of digital content and access to it, providing guarantees on the quality and integrity of curated data and content, implementing access control policies which are compliant with the protection of intellectual property. DOI, URN, ARK, and PURL are established schemes for access to objects, whilst Scopus Author ID, Researcher ID and ORCID belong all to this approach for people information.

An alternative approach, called the **Linked Data approach**, has more recently emerged from the Semantic Web and "open data" communities. This approach was not started as an approach to persistent identifiers, but rather as a way of enabling the creation a **web of data** (a global space of interlinked datasets), which has the potential to reproduce the network effect that the web of documents had on hypertexts. While there is a concern within the PID community about the possibility that, through Linked Data, the web itself may be taken as the platform for e-Science and the current practices about HTTP URIs as a way of managing persistent identifiers, there is also an increasing awareness in part of the PID community that the Linked Data practices and tools may offer an opportunity for extending the value of data (in particular, through cross-linking) and cover use cases which traditional solutions were not designed to address (for example, identifying dynamic resources).

The opposition between the two approaches shows that, if some technical challenges need to be addresses to encourage and motivate an agreement among stakeholders on converging towards commonality in their solutions, technology is not the main driver in leading this process. Any identifier solution is always used within cultural, geographical, disciplinary and organizational boundaries through a technical system and the process of reaching an agreement between parties over possibly conflicting purposes and objectives is a process which is played out at the interfaces of these boundaries.

A multidimensional analysis

From these premises, it follows that devising an appropriate solution to the problem of identifiers is far from being a marginal or merely technical issue in the design of SDIs. The role of identifiers in addressing issues like data quality assessment, access, preservation and provenance goes far

> *"The role of identifiers […] goes far beyond the technical level to embrace a much wider vision, where organizational, social and business strategies form an intertwined eco-system."*

beyond the technical level to embrace a much wider horizon, where organizational, social and business strategies must be taken into account in considering valuable ways forward. Another important dimension, which contributes to the complexity of the SDI landscape deals with the key players, which populate the eco-system and influence, with their needs, expectations and requirements, its internal dynamics.

There is also a temporal dimension at play, as issues and corresponding solutions are currently at different stages of awareness and maturity, and sustainable models must take this heterogeneity into account. This is confirmed by the results of a recent survey, which has been conducted in the context of the EU co-funded APARSEN network of excellence about the use of persistent identifiers (PIs) for digital objects and authors: 86% of the respondents reported to use PIs for digital objects and only 47% to use PIs for authors. This evidence confirms that there is a different level of maturity between the more advanced systems for digital objects and the gradually emerging solutions for authors.

It follows that the use of identifiers operating across boundaries represents a multidimensional challenge, which involves many layers of interoperability and requires co-ordination of activities and responsibilities across these boundaries.

If a holistic stance towards SDIs is taken, incorporating technical, social, geographical and organisational factors, defined boundaries of responsibility begin to emerge between localised environments involving identifiers. Evidence suggests there is opportunity to act at the interfaces between these localised activities, in order to maximise value extracted through enabling interoperability between the localised environments.
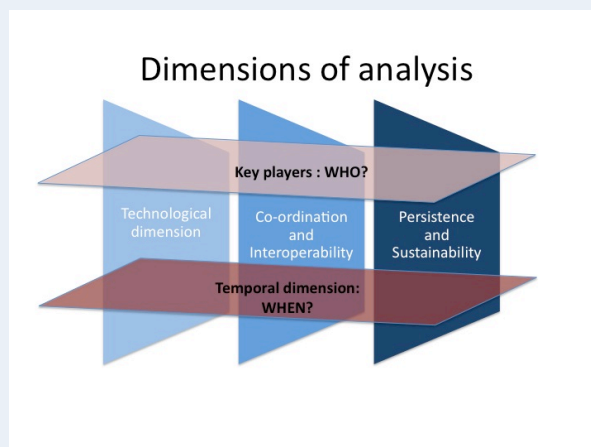
In order to provide a comprehensive picture of the complex landscape in which SDIs and identifier systems are being developed, the approach adopted in DIGOIDUNA was to perform a **multi-dimensional analysis** based on three main dimensions of investigation. The first dimension deals with **technological aspects**, the second dimension incorporates **co-ordination and interoperability factors** and, finally, the third dimension concerns with **persistence and sustainability issues**. The analysis includes two other dimensions, orthogonal to the previous ones, which deal with the relevant **key players** involved in the process under investigation and the **temporal evolution** of the process itself. The combination of all these dimensions reveals a complex eco-system where many levels of interoperability need to be promoted through coordination actions in order to increase the potential of innovation and the advancement of the research in SDIs.

These points illustrate that beneath identifier schemes and their attendant management systems there must be agreement at many different levels: consensus between the parties involved in their creation and use, over the operational responsibilities (service level agreements), and allocation of costs involved. Agreement is also needed as to what the identifiers should stand for, represent, and 'mean'[3]. The *process* of reaching consensus between parties over possibly conflicting purposes and objectives is clearly not a technical concern (although agreements over technical issues are required), but is a critical one.

In summary, any solution to the problem of identifiers for digital objects and authors requires to embrace a much wider vision, where organizational, social and sustainability strategies form an intertwined eco-system which must considered in devising solutions for identifiers in SDIs. This multidimensional perspective is fundamental to make a diagnosis of the current situation and propose integrated actions toward co-ordinated sustainable and agreed identifier solutions to promote the potential of SDIs in enabling innovation and scientific advances.

3   Digital Identifiers: the gap to be filled

As it emerges in the first part of the report, there is a gap between where we are (the current situation) and the desirable services and scenarios which would be enabled if a full-fledged solution for managing identifiers for digital and non-digital objects in SDIs was in place (as illustrated in the scenarios described in Section 2). In this section, the gap is diagnosed using SWOT analysis as a methodological tool to carry out a multidimensional analysis. On the one hand, the output of the analysis can be used to identify the strengths upon which policy makers can build an action plan to move forward, together with the opportunities which can be exploited to make action more effective; on the other hand, it can be used to raise awareness about the weaknesses of the situation in which action must be taken and the threats which might jeopardize the successful outcome of action.

The outputs of the diagnosis are summarized in the table below and can be described as follows:

Strengths

1. In the last few years, Europe has made significant investments towards promoting e-Science as a new way of conducting research, and in making available Scientific SDIs for enhancing and making more effective the creation, sharing and reuse of data collected by teams of researchers in different geographical locations. These investments triggered deep changes on the process of scientific production, enabled new methods of research (e.g. *in silico* experiments) and made possible the creation of active virtual scientific communities in every discipline. The road to e-Science is paved, and no one can seriously think that scientific production will go back to the previous situation. But this in turn leads to a strong need to demonstrate to the public opinion that the investments made in e-Science and SDIs have an adequate return on investment. The experts in the field seem to agree in the fact that a suitable infrastructure for managing DIs is one of the key elements for maximizing the "Return-On-Investments" (ROI) from e-Infrastructures, as DIs are the key enablers of value-added services on top of scientific data and scholarly content.

---

[3] For example, via explicit standards and documentation or via machine-based techniques as used on the semantic web.

2. One of the key findings of this study is that the know-how needed to design and deploy a suitable infrastructure for managing DIs is available today. Among experts, there is a wide agreement that the toughest challenges in delivering such an infrastructure are not at the technical level (even though there are some open technical issues which must be addressed as RTD work), but at the level of creating a common conceptual framework and a common agenda among key players and stakeholders. Building new technical skills from scratch in new areas may require several years, and success is not always guaranteed. But, for the problem under scrutiny in this study, the necessary technical skills and know-how are not a significant issue.

3. In the last decades, several initiatives of different size, breath and governance structures have been started to address the problem of managing identifiers for digital objects and authors. Beyond initiatives, which aim at providing *ad hoc* solutions within local systems, some pioneering efforts are underway aiming at providing solutions, which go beyond the boundaries of a single institutional or commercial system and address the issue from a cross-boundary perspective. These integrating initiatives, among which we mention DOI, Handle, URN, ARK, PURL for digital objects and gradually emerging solutions like ORCID for authors, pose crucial challenges and opportunities for the future of SDIs in terms of layers of interoperability, social agreements on common policies, integration of different sustainability models.

Weaknesses

1. The main weakness is the lack of consensus and coordination among the parties involved in establishing an open, dynamic and sustainable governance of e-infrastructure for e-Science using identifier systems for digital objects and authors. This has led to a still fragmented environment where solutions are still orchestrated by the needs and requirements of few parties and conflicting interests tend to hinder the adoption of common policies on the governance of identifiers for digital objects and authors.

2. From a financial and business perspective, the situation is still immature. Financial sustainability is only recently becoming an issue in the digital identifier ecosystem and traditional funding schemes appear to be inadequate to address sustainable solutions. This is confirmed by the results of the already mentioned survey conducted in the context of the APARSEN project about PIs for authors, digital objects and organizations. The results show that there is a disparity between the reported need of adopting some system for managing persistent identifiers for their digital resources (claimed by more than 60% of participants) and the scarce commitment in terms of financial sustainability. Around 50% of participants reported a lack of a specific funding model for digital preservation practices (including persistent identifiers). The lack of diversification and flexibility of funding models is a further impediment for the development of scaling and promising solutions that embed or promote the value of identifiers in SDIs.

3. There is a scarce level of awareness within many stakeholder communities about the benefits of DIs for the creation of value-added services on top of scientific data and content. This has important implications in terms of adoption and the contribution by the relevant stakeholders in the participative definitions of governance structures and policies.

Opportunities

1. An important opportunity in the context of establishing SDIs deals with the role of identifiers in the consolidation of e-Science. The rising tide of scientific and non-scientific digital data has encouraged the development of e-Science infrastructures in laying the basis for modern science. In this context, the possibility of accessing an amazing amount of scientific content and data in digital format, the emergence of new working research methods enabling pan-European and worldwide collaborations between researchers, the increasing linkage across articles and datasets, the development of new and much more powerful metrics for assessing impact of scientific production are great opportunities which however should not hide the emergence of new challenges, like for example digital preservation, quality assessment and certification. Identifier systems play a crucial role in making SDIs competitive in the face of these challenges and foster scientific progress and innovation.

2. Another opportunity deals with science-based innovation in the future. By facilitating data and knowledge sharing across sectors and national borders, SDIs – through DIs – may have a crucial role in helping the transition of novel technologies and services to the marketplace and transforming them into goods with economic and social value. The transfer of knowledge and expertise beyond scientific areas has the potential to create new synergies between private/commercial sectors and the scientific arena, opening new frontiers to scientific discovery.

3. Empowering the cross-discipline fertilization and knowledge sharing between academia and industry, DI systems open a big opportunity in terms of ROI for public interventions by increasing the potential of innovation. Funding schemes may include specific activities to reinforce the partnership with industry, realizing the potential to produce socio-economic value, which spans a diversity of sectors, from cultural industries to manufacturing.

4. Despite the level of immaturity in terms of polices, business models and sustainability, accumulated experience has provided a common ground to set and find viable technical foundations for DI. These technical foundations represent a crucial asset to move forward toward integrated and cross-boundary solutions that will have a long-lasting impact on scientific research and innovation in Europe and worldwide. The opportunity here is to optimize existing solutions and technologies through integrating activities.

Threats

1. The change of priorities in the political agenda or in funding represents a potential element of fragility, which may have an impact on the financial sustainability and scaling of agreed solutions.

2. The lack of institutional consensus on strategic directions is a second element of vulnerability, which can hinder the effectiveness and coherence of national and EU policies and international cooperation in the field of SDIs.

3. The resistance to change within the scientific community, accompanied by a misconception about the potential impacts of the usage of DIs within SDIs and related services (e.g. evaluation metrics), represents an important threat for reaching common policies and social agreement between parties.
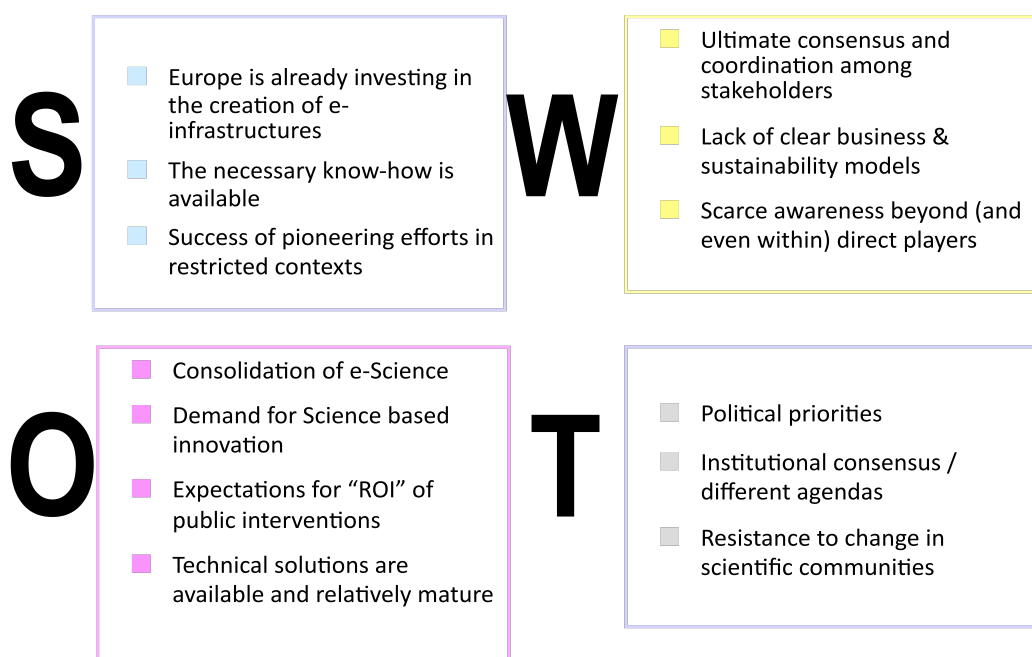
S
- Europe is already investing in the creation of e-infrastructures
- The necessary know-how is available
- Success of pioneering efforts in restricted contexts

W
- Ultimate consensus and coordination among stakeholders
- Lack of clear business & sustainability models
- Scarce awareness beyond (and even within) direct players

O
- Consolidation of e-Science
- Demand for Science based innovation
- Expectations for "ROI" of public interventions
- Technical solutions are available and relatively mature

T
- Political priorities
- Institutional consensus / different agendas
- Resistance to change in scientific communities

Figure 3: Summary of the SWOT analysis

4 Digital Identifiers: actions needed to fill the gap

Based on the SWOT analysis reported in the previous section, the final part of this report focuses on a collection of actions which can help filling the gap by capitalising on connections and commitments between stakeholders where agreements exist, engaging new initiatives and developments where necessary, and working together to ensure a coordinated ecosystem of identifiers can be built into the heart of a coordinated infrastructure, in order to ultimately maximise the multidimensional value from supporting all aspects of scientific production. The underlying rationale for recommendations is that:

- in order to foster the growth of DI-based solutions, appropriate measures are needed in different fields of action, which are depicted in Figure 4. These measures must aim at(1) defining new governance models, (2) raising awareness and social support among potential users, (3) fostering demand of DI-based solution and services, (4) redefining DI-oriented business ideas and (5) finally promoting RTD and standardization activities;
- support measures must be coordinated within an "agreed upon" portfolio approach, securing complementarities/synergies and whilst achieving efficiencies by operating at the boundaries of localised domains, accepting a certain level of overlap in the implementation of such measures.

The recommendations are also structured around four main macro-categories (see Figure 5). Each category corresponds to a single meta-recommendation, which is structured around a small set of specific actions. The central category deals with the definition of a conceptual framework for DI management, which embraces, at a conceptual level, aspects that are implemented by the other three categories. They include: 1) bootstrapping DIs in SDIs, 2) mobilizing resources and 3) defining sustainable solutions. In the following part of this section, each category is presented in details with the corresponding recommendations. An overview of the recommendations, classified by colour according to the above four dimensions and their relationships to the potential areas of intervention described in Figure 4, is shown in Table 1. The potential impact of the recommendations and risks of inactions are also reported.
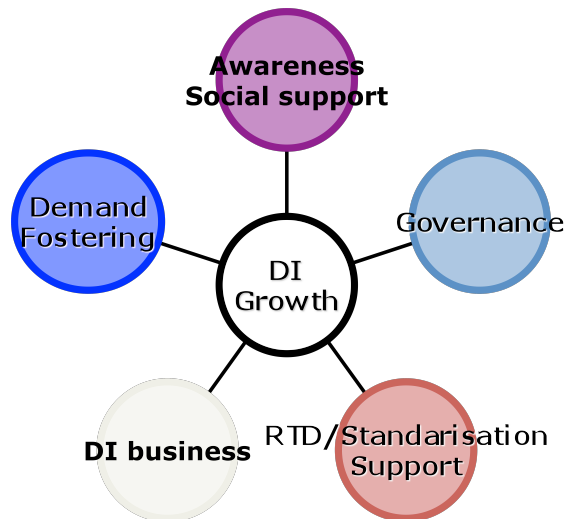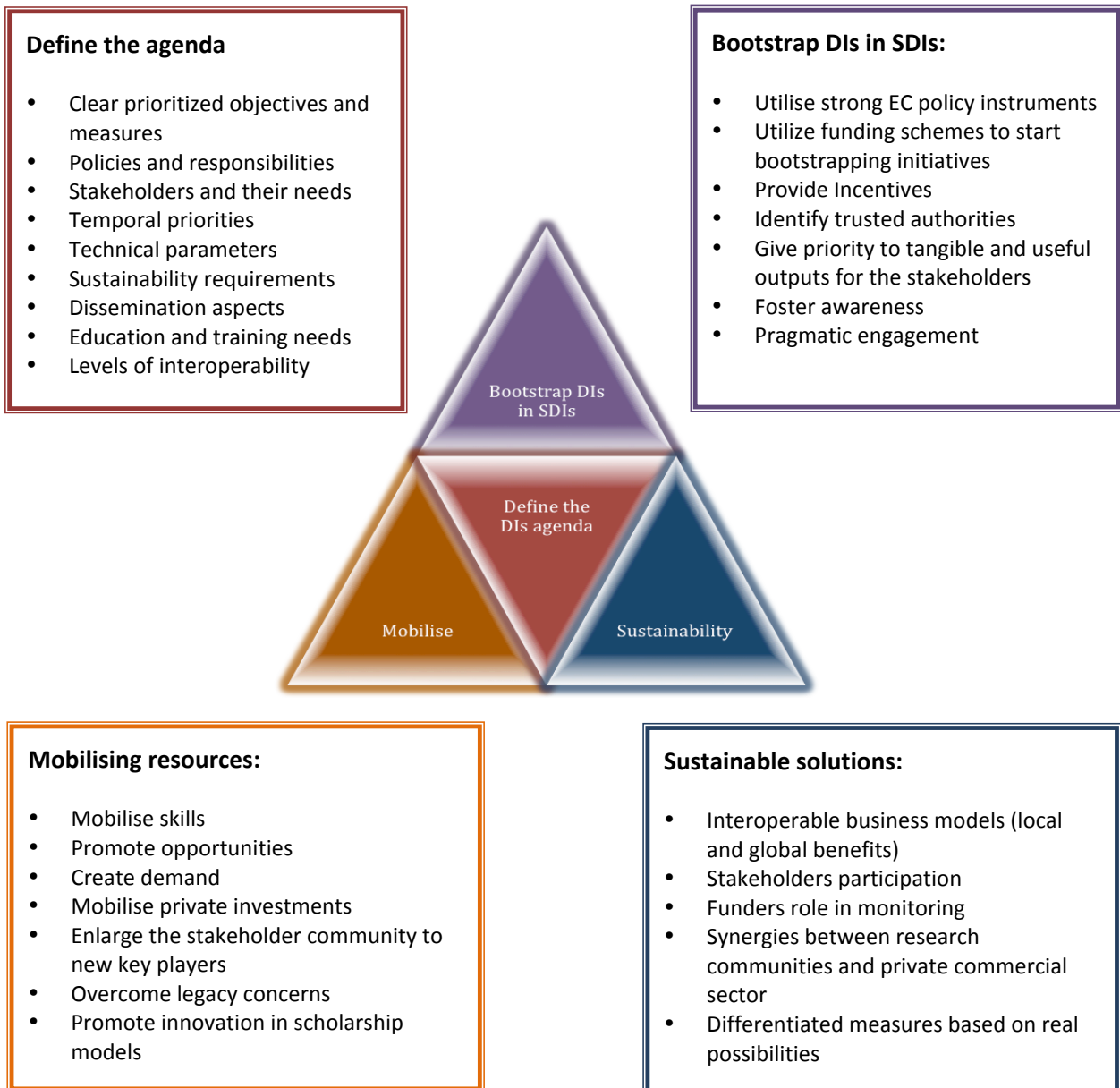
Figure 4: Fields of Action

Define the agenda

- Clear prioritized objectives and measures
- Policies and responsibilities
- Stakeholders and their needs
- Temporal priorities
- Technical parameters
- Sustainability requirements
- Dissemination aspects
- Education and training needs
- Levels of interoperability

Bootstrap DIs in SDIs:

- Utilise strong EC policy instruments
- Utilize funding schemes to start bootstrapping initiatives
- Provide Incentives
- Identify trusted authorities
- Give priority to tangible and useful outputs for the stakeholders
- Foster awareness
- Pragmatic engagement

Mobilising resources:

- Mobilise skills
- Promote opportunities
- Create demand
- Mobilise private investments
- Enlarge the stakeholder community to new key players
- Overcome legacy concerns
- Promote innovation in scholarship models

Sustainable solutions:

- Interoperable business models (local and global benefits)
- Stakeholders participation
- Funders role in monitoring
- Synergies between research communities and private commercial sector
- Differentiated measures based on real possibilities

Figure 5: macro-categories of recommendations

16

Define the DIs agenda

The EC should start initiatives aiming at defining a common agenda among key stakeholders towards the design and implementation of a governance model and an integrating infrastructure for managing DIs in SDIs in which technological, economical, social and political factors are taken into account.

Given the currently complex landscape of DIs across SDIs, a common direction must be defined. To realise this effectively, the EC should capitalise on its connections and seek commitments between stakeholders where agreements exist, engaging new initiatives and developments where necessary, and ensure parties work together to ensure a coordinated ecosystem of identifiers can be built into the heart of a coordinated infrastructure from the outset. To this purpose, a common DI agenda should be defined and agreed by the key stakeholders. The agenda will define a clear conceptual framework, which will be a pre-requisite for dialogue and achieving consensus across the communities impacted, and serving as the basis for promoting awareness and mobilisation of skills and resources.

The common agenda should cover at least the following issues:

1. Defining the common objectives and organize them into a list of workable temporal priorities.
2. Agreeing on a shared governance model, which defines devolved responsibilities amongst stakeholders and ensures long-term sustainability.
3. Sharing a conceptual framework in which the basic technical parameters and the fundamental services are introduced and described.
4. Planning interventions to promote awareness, dissemination and education activities aiming at expanding and reinforcing DI knowledge and skills.

Bootstrapping DIs in SDIs

The EC, Member States and other relevant stakeholders must take specific actions aiming at bootstrapping the implementation of the DI agenda in order to secure a critical mass of coordinated DI systems.

The fragmentation of the current landscape and the different needs / interests of the stakeholders involved in different initiatives may lead to a situation which can hinder the actual deployment of an integrated DII for SDIs. Therefore, actions are needed to bootstrap the convergence toward common policies on the governance of identifiers and toward integrating technical solutions. To effect change, the EC should utilise strong policy instruments, and the outputs of initiatives receiving EC support must be tangible and usefully contribute to the DII goals through being valuable – i.e. of direct use – to SDI stakeholders. In particular, they must drive demand, foster trust, and allow for measurement, reassessment and therefore for potential adjustment of policy through experience gained and lessons learned. In addition, these actions should work towards systematic implementation of those technical and organisational factors that underpin trust in identifiers and the services built upon them – thereby encouraging the perception and reality of reliability as a key component of SDIs. Where possible, consolidation of established systems seems appropriate rather than creation of new ones or an increased proliferation of those localised solutions unable to leverage centrally-managed services and resources.

Factors affecting priority of bootstrapping actions should include likelihood and impact assessment of near-term gains (the "low hanging fruit"), value of returns, the difficulty and relevance of challenge to demonstrating value of DIs. Given the pervasive nature of effective usage of DIs anticipated, early-funded solutions should operate similarly cross-scale: locally, at institutional level, at national infrastructure level,

at EU level, and proactively coordinate within the global research infrastructure. Solutions should technically seek to interoperate across all levels - across disciplinary, national/geographical, and community boundaries.

Some specific exemplary actions that encapsulate the recommendations above are:

1. Reinforce, promote and secure EU wide institutional coordination among vertical and regional clusters of e-infrastructure stakeholders (common policies on the governance of identifiers for digital objects and authors).
2. Funding bodies must provide initial support to seed initiatives, which aim at implementing the coordination model defined in the agenda and at creating a critical mass of coordinated DI systems. This must be done in a flexible way, which allows the reallocation of funds in the portfolio based on the emerging needs and requirements.
3. Promote awareness and skills development to enable different stakeholders to participate effectively on DI initiatives and infrastructures.
4. Work towards systematic implementation of technical and organisational factors that underpin trust in identifiers, their reliability as a key component of SDIs - secure their operational management.

Mobilizing Resources

Stakeholders at any levels should promote actions to mobilize technical, human, financial resources aiming at triggering a wider demand of usage and exploitation of e-Science results based on DIs.

Deploying productive and competitive DI solutions for SDIs requires contributing to the mobilisation of technical, human and financial resources. The underlying rationale is that the Europe has a vast skills pool needed to implement DI-based added value services and is well-positioned to take on the challenges required to implement digital identifiers infrastructures (DIIs) effectively within SDIs. A first example is the mobilization of stakeholders in the public and private sectors to co-operate, for example by implementing public-private partnerships under national DI strategies as well as through international collaboration schemes to address global scientific challenges. There is a need to reinforce the partnership with industry and other pioneers from institutional bodies, aimed at ensuring the long term preservation of digital assets in their respective fields of competence but also fostering the use of scientific infrastructures beyond the scientific area.

The stakeholder community must be enlarged to include new key actors and create wider demand, to implement long-term sustainability and promote economic viability and robustness. However, institutional stakeholders must be incentivised to develop their solutions to take account of this to discourage ad-hoc solutions, promote reuse over reinvention, adopt methods to ease identification of digital entities (authors and objects), interconnecting with existing systems provisions should it not be possible to build in solutions directly. These activities should be regarded as opportunities to overcome legacy concerns, to therefore mitigate risk of "islands of non-interoperability" persisting further – although care will need to be taken to ensure where areas of transition must be navigate carefully and ensure overall that the business needs of the sector does not become disrupted, a concern echoing current (and related) discussions over open access to journal publications. Actions instead should serve to promote innovation in scholarship models, so that scholarship serves the wider European knowledge economy to the maximum effect possible. Private investments should be mobilised wherever possible; the future EC Framework Programme Horizon 2020 should play a catalysing and leveraging role. Measures defined to assess the quality of DI systems can be used to mobilise specific resources.

Participation should be promoted across the spectrum: from individual researchers "claiming", verifying or enhancing their profile data, the various roles and responsibilities institutions may adopt (e.g. as publishers of research outputs, knowledge producers and consumers, preservation and curation services providers, authorities of information about entities, sources of teaching and learning materials), to funding bodies (providers and consumers of analytic data), commercial interests, and society as a whole (through e.g. citizen science, general media interest and scrutiny of funding allocation).

Some specific examples of actions to this end are:

1. Funding agencies should design funding schemes, which may attract new public and private investments and efforts in developing and adopting DI-based added value services and solutions.
2. Stakeholders, and especially funding agencies, should foster interoperability based on consolidation of established DI systems (where possible) more than on proliferation of *ad hoc* systems.
3. Actions should be taken to mobilize consolidated technical skills (including those deriving from the web of data community) to implement effective digital identifiers infrastructures (DIIs) within SDIs and adopt measures to assess the quality and impact of them for the exploitation of e-Science results.

Sustainable Solutions

Efforts should be invested to build suitable organisational mechanisms and business models to guarantee the-long term sustainability of DI solutions.

A key factor which promotes trust and adoption is the long term sustainability of the infrastructure for managing DIs. Therefore, sustainability goals and measures for monitoring and reporting progress must be defined at the outset within the overall framework. Wide stakeholder participation is key to business agility, resilience and robust viability under varying economic and global conditions. Funders have a major role in monitoring the effectiveness of interoperability between communities and commercial practice, and to prevent silos developing that might serve to damage the interests of the ecosystems served by SDIs as a whole. Funding bodies should also strive to develop opportunities between scientific research production and the commercial sector, in terms of collaborative actions and models, creation of synergies, creation of opportunities for exchange ("DI-related Public-Private Partnerships" -PPP-).

Stakeholders need to develop business models that serve their own interests, yet are interoperable; which can deliver local benefits and at the same time contribute to global effects, as the nature of the value of DIs accumulates through usage, as well as accuracy and longevity and other curation services. Resilience should be built-in, implying clear allocation of operational responsibilities with appropriate justification of cost models, where the cost of infrastructure is offset amongst the beneficiaries of the services provided (within SDIs as a whole).

Some specific actions that encapsulate the commentary above are:

1. Stakeholders need to develop business models where the costs of developing and sustaining identifier infrastructures and the responsibility in granting the long term sustainability of these infrastructures are distributed among the beneficiaries.
2. The flexibility of funding sources should be enhanced, allowing the reallocation of funds in the portfolio to enable the rapid scaling of promising solutions that embed or promote the value (usage) of identifiers.
3. Funding bodies must support the development of collaborative models and actions to create synergies and exchange opportunities between the private/commercial sector and scientific sector -DI- PPP).

Conclusions

Identifiers lie at the crux of the issues examined by the study – they are the means to manage not only data within SDIs, but the control and supply of knowledge that delivers value to all stakeholders within research production. Whilst there are risks and drawbacks associated with forward-looking changes, at the same time the possibility of a coherent overall approach towards identifiers presents tremendous opportunities for extracting value from the accumulated knowledge and outputs of research production across the European research landscape.

This Report summarises the study's recommendations and their justification. The recommendations at their heart form a set of proposals for actions to be taken not only at European and Member State level, but involving all stakeholders, to develop an open and sustainable e-Infrastructure for locators of digital objects and identifiers of authors supporting scientific information access, curation and preservation.

Recommendations were organised along two distinct trajectories.
- The first trajectory is in a way that takes account of the lifecycle of the elements in an ecosystem that addresses the issues within the scope of the study. The trajectory is organized in four key actions: defining a common agenda, bootstrapping a critical mass of coordinated DI systems, mobilising resources and ensuring sustainability.
- The second trajectory is organised in terms of measures aimed at addressing specific drivers, *viz.* creating a critical mass of coordinated actions, securing long-term viability, creating demand and the conditions for exploitation, and guidance for technical standardisation and RTD-related activity.

The basic argument of this report can be summarized as follows. First of all, *digital identifiers are the key to creating value from SDIs and their cost effective management.* Second, *moving from a local DI governance and management models to cross-boundary interconnected solutions poses new challenges for managing identifiers*. Third, *the role of identifiers goes far beyond the technical level to embrace a much wider vision, where organizational, social and business strategies form an intertwined eco-system*. The recommendations and proposed actions must be read in this context as a way of bridging the gap between the current situation and a future scenario in which DIs become a crucial source of value for SDIs.

Summary recommendation	Specific Actions	Areas of Intervention	Potential Impact	Risk of Inaction
Define the DIs agenda				
The EC should start initiatives aiming at defining a common agenda among key stakeholders towards the design and implementation of a governance model and an integrating infrastructure for managing DIs in SDIs in which technological, economical, social and political factors are taken into account.	1) Defining the common objectives and organize them into a list of workable temporal priorities. 2) Agreeing on a shared governance model which defines devolved responsibilities amongst stakeholders and ensures long-term sustainability. 3) Sharing a conceptual framework in which the basic technical parameters and the fundamental services are introduced and described. 4) Planning interventions to promote awareness, dissemination and education activities aiming at expanding and reinforcing DI knowledge and skills.	Governance	• Movements towards agreements across parties • Coordination actions will be more effective embracing all the dimensions of the complex eco-system of SDIs • Consensus between the key stakeholders over what features constitute *value* within an identifier scheme/system • Adoption of common measures	• Fragmentation • Lack of coordination • Proliferations of legacy systems
Bootstrap DIs in SDIs				

The EC, Member States and other relevant stakeholders must take specific actions aiming at bootstrapping the implementation of the DI agenda in order to secure a critical mass of coordinated DI systems.	1) Reinforce, promote and secure EU wide institutional coordination among vertical and regional clusters of e-infrastructure stakeholders (common policies on the governance of identifiers for digital objects and authors).	Governance	• Coordination • More participative governance structure • Harmonization of interests of different stakeholders • Stakeholders contribute to the definition of the requirements of PIs systems and this participation increase the adoption of these systems. • Shifting from technology driven development to stakeholder-needs driven development of ID systems • Capitalising on Europe's strong available of skills and expertise to implement model within technology • Stakeholders' trust and reliability of the solutions • Overcoming the new challenges to	• Proliferation of local and ad hoc solutions • Lack of support or interest amongst stakeholder groups in reaching consensus • Lack of awareness and low adoption • Lack of trust in DI systems
	2) Funding bodies must provide initial support to seed initiatives which aim at implementing the coordination model defined in the agenda and at creating a critical mass of coordinated DI systems. This must be done in a flexible way, which allows the reallocation of funds in the portfolio based on the emerging needs and requirements.	RTD Demand fostering		
	3) Promote awareness and skills development to enable different stakeholders to participate effectively on DI	Awareness and social support		

	initiatives and infrastructures.		the protection of personal data and privacy	
	4) Work towards systematic implementation of technical and organisational factors that underpin trust in identifiers, their reliability as a key component of SDIs - secure their operational management.	RTD Governance	• Stakeholders confidence and satisfaction increase • The development of ID systems and SDIs in not seen as a technical issue but as a multidimensional matter -> more holistic approach • Converting the intellectual advantage of research into competitive advantage of innovative solutions	
Mobilize resources				
Stakeholders at any levels should promote actions to mobilize technical, human, financial resources aiming at triggering a wider demand of usage and exploitation of e-Science results based on DIs.	1) Funding agencies should design funding schemes, which may attract new public and private investments and efforts in developing and adopting DI-based added value services and solutions. 2) Stakeholders, and especially funding agencies,	DI business Demand fostering Awareness/social support Demand fostering	• Support for data producers • Improving the access to publicly funded data • Stakeholder engagement increase • Unlocking the potential of IDs through added value services • Solutions incorporate not replace current ones	• Barriers that may hinder adoption of identifier solutions • Development of low-value services • Proliferation of local solutions • Lack of interoperability with emerging approaches to manage identifiers for digital objects

	should foster interoperability based on consolidation of established DI systems (where possible) more than on proliferation of *ad hoc* systems. 3) Actions should be taken to mobilize consolidated technical skills (including those deriving from the web of data community) to implement effective digital identifiers infrastructures (DIIs) within SDIs and adopt measures to assess the quality and impact of them for the exploitation of e-Science results.	RTD/ Standardization support Demand fostering	• bootstrapping the web of data • lowering the barriers to publishing and accessing data on the Web	and authors (e.g. Linked data community)
Sustainability				
Efforts should be invested to build suitable organisational mechanisms and business models to guarantee the-long term sustainability of DI solutions.	1) Stakeholders need to develop business models where the costs of developing and sustaining identifier infrastructures and the responsibility in granting the long-term sustainability of these infrastructures are	DI business	• Long-term sustainability and sharing of responsibilities between key stakeholders • Stimulating diversified investments complemented by targeted public investments to	• Difficult to prioritise funding, identify cost-saving measures and find innovate models within research production due to lack of measurable outputs; especially in

			improve resource allocation and sustainability	terms of commonly-accepted metrics.
	2) The flexibility of funding sources should be enhanced, allowing the reallocation of funds in the portfolio to enable the rapid scaling of promising solutions that embed or promote the value (usage) of identifiers.	DI business	• Long-term sustainability • Leverage more private investment • Coordinated allocation and re-allocation of the resources • Cross-fertilization between scientific and private/commercial sectors, busting the number and impact of collaborative efforts between the two sectors • New public-private partnerships can be stimulated • Leverage more private investment increase the knowledge sharing between academic and industrial researchers and strengthen the base knowledge of industry • Research fuels the knowledge economy	• The effective use of identifiers can be inhibited by silos of responsibility • Lack of re-allocation of financial resources • Lack of scaling of financial solutions • Misalignment between private and academic/scientific sectors in terms of needs, expectations, objectives and so on. • Lack of coordination and pooling of resources
	3) Funding bodies must support the development of collaborative models and actions to create synergies and exchange opportunities between the private/commercial sector and scientific sector ("DI-related PPPs").	Governance Demand fostering		

Table 1: An overview of the DIGOIDUNA recommendations

Expert working group

Paolo Bouquet, *leading author* (associate professor, University of Trento, Italy)

Barbara Bazzanella (researcher, University of Trento, Italy)

Andrea Molinari (University of Trento, Italy)

Martin Dow (expert, Acuity Unlimited Ltd., UK)

Ruben Riestra (expert, Spain)

References

[COM2009] ICT Infrastructures for e-Science. Communication from the European Commission, 2009.

[COM2010] A Digital Agenda for Europe. Communication from the European Commission, 2010. http://ec.europa.eu/digital-agenda/

[FP2009] Hey, Tony; Stewart Tansley and Kristin Tolle, Eds. "The Fourth Paradigm: Data-Intensive Scientific Discovery." Microsoft Research. Redmond, Wash: 2009. Available as PDF at http://research.microsoft.com/enus/collaboration/fourthparadigm/

[GRDI2020] GRDI Preliminary Roadmap Report: "*Global Research Data Infrastructures: The Big Data Challenges*". http://www.grdi2020.eu/Pages/SelectedDocument.aspx?id_documento=533f71bd-a415-4dfe-af67-b4a43f14a755

[IPI2006] Hans-Werner Hilse and Jochen Kothe, ``*Implementing persistent identifiers*", 2006. http://pid.ndk.cz/dokumenty/zakladni-literatura/urn_nbn_de_gbv_7-isbn-90-6984-508-3-8.pdf

[ISTAG2011] Orientations for EU ICT R&D & Innovation beyond 2013 - 10 Key Recommendations - Objectives, rationale and content". Report from the Information Society Technologies Advisory Group (ISTAG), 2011. http://cordis.europa.eu/fp7/ict/istag/documents/istag_key_recommendations_beyond_2013_full.pdf

[KT2002] Kenneth Thibodeau, *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*. Report of the Council on Library and Information Resources (CLIR). http://www.clir.org/pubs/reports/pub107/thibodeau.html

[Nat2008] Jane Qiu, *Scientific publishing: identity crisis. Nature*. 2008 Mar ;451(7180):766-7. DOI: 10.1038/451766a

[PersID2011] "Current State and State of the Art and User Requirements". PersID project report, http://www.persid.org/downloads/finalreports/PersID_Report_Part_3_final.pdf

[RtW2010] "Riding the wave - How Europe can gain from the rising tide of scientific data". Final report of the High level Expert Group on Scientific Data. A submission to the European Commission. October 2010. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

 [SDI2009] Kostas Glinos. "Scientific Data e-Infrastructures – Supporting the transition to e-Science". Presentation at the Alliance for Permanent Access Conference. 24 November 2009, The Hague. http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/02/08_KonstatinosGlinos_EC_241109.pdf

 [TAG2004] *Architecture of the World Wide Web*, Volume One. W3C Recommendation 15 December 2004. http://www.w3.org/TR/2004/REC-webarch-20041215/