GEDE PID Focus Area Preparation

Peter Wittenburg

28. 1. 2017

This document is meant to prepare the discussions of the Focus Group on PIDs. We suggest dealing with the following questions:

1. Identify which other relevant initiatives[[1]](#footnote-1) made assertions important to the PID Focus Area.
2. Clarify all assertions and statements making sure that there is a common understanding.
3. Finding agreement and disagreement.

In the coming virtual meeting and in its preparation we should first focus on points 1 and 2 and leave point 3 to the next meetings. We need to make sure that our terminology is harmonised far enough.

We can identify the following contributions until now:

* The following grouped list of assertions (see appendix 1 which has been extended based on the original list in appendix 2) has contributions from
	+ RDA DFT WG
	+ FAIR Principles
	+ RDA PIT WG
	+ RDA DFIG IG discussions
	+ DOI - IDF
	+ PID Workshop in Munich
	+ ITU
* At the GEDE meeting in Bratislava the following issues were raised (see appendix 3)
* the related issues of granularity and collection building
* how to do versioning
* which kind of metadata to use in PID records
* binding role of PIDs
* semantic categories.
* when to assign PIDs

The following procedure is being suggested:

* First Focus Group virtual meeting at 1.2. focussing on points 1 and 2
* Second Focus Group virtual meeting at begin of March focussing on point 3 and the list of assertions
* Plenary session in Barcelona focusing on point 3 and discussing the GEDE meeting usage aspects.

Grouped List of Assertions

**Nature of PIDs and PID Systems**

**PID2. RDA DFT1.2:** A persistent identifier is a long-lasting ID represented by a string that uniquely identifies a DO and that is intended to be persistently resolved to meaningful state information about the identified DO.

**PID3. RDA DFT1.3:** A PID record contains a set of attributes stored with a PID describing DO properties.

**PID4. RDA DFT1.4:** A PID resolution system is a globally available infrastructure system that has the capability to resolve a PID into useful, current state information describing the properties of a DO. (*state information can be interpreted as systems metadata*)

**PID8. RDA**: A trustworthy PID system must

* be maintained by a dedicated and reliable team,
* be based on a transparent sustainable business model,
* be provided by a non-profit organisation,
* be subject of regular quality assessments by external parties,
* be governed by international boards,
* be based on open standards,
* be based on a redundant and secure architecture,
* support a huge address space (*comparable or even larger than IPv6*) and
* support an openly documented API optimally supporting accepted data models.

**PID 25. ITU**: The platform for interoperability of heterogeneous identity management systems is an open architecture based on ITU-T X.1255 and the Digital Object Architecture ... and is capable of to offer interoperability at global level.

**PID 26. ITU:** The top 5 benefits from this platform are: 1) framework to enable interconnection of objects, data, devices and processes, 2) in-built security regime (PKI) and data privilege delegation, 3) multilingual support and access to a variety of type value pairs, 4) enable defining new type value pairs for increased flexibility for new types of services and applications, 5) interoperable with existing identity management systems.

**Relevance of PIDs and PID Systems**

**PID 18. PID WS**: International and national steps need to be taken urgently to offer a sustainable, structured and mature PID service landscape based on quality assessed service providers to all interested parties. Only such a structured and massive approach will prevent ending up with unresolvable PID zombies.

**PID 19. PID WS**: PIDs are becoming essential across sectors and communities for different application scenarios and efforts need to be taken to offer services across these sectors and communities.

**Assigning PIDs**

**PID1. RDA DFT1.1:** A digital object is ... referenced and identified by a persistent identifier ...

**PID5. FAIR-F1**: (meta)data are assigned a globally unique and eternally persistent identifier

**PID11. RDA**: A PID needs to be requested as early as possible, at least at the time of registration at a trustworthy repository a PID record needs to be available.

**PID12. RDA:** PIDs are associated with collections which can exist of a number of digital entities, i.e. the level of granularity at which PIDs will be assigned is left to the communities and repositories. A high granularity is recommended to anticipate future applications.

**PID16. PID WS**: Proper PID usage and support will become key for competitiveness in science and industry.

**Using PIDs**

**PID6. FAIR-A1**: (meta)data are retrievable by their identifier ...

**PID7. RDA-PIT1**: PID systems should support the generic PIT API where Information Types (*properties of DOs)* are openly registered and defined.

***This assertion needs to be adapted to: PID systems should provide the attribute profile they are supporting under their prefix root. The API was a demo setup and is replaced now.***

**PID9. RDA:** The PID Record can be used to store the context of digital objects (bitstream locations, metadata, PID, rights information, landing page, etc.)

**PID13. RDA:** A metadata description contains the PID of the corresponding object. The PID record contains the metadata PID to ensure at all times that DO's context can be retrieved. (this can be compared with the reverse DNS mechanism)

**PID15. RDA**: The PID record should include an expiration date for the digital object. Even for digital objects that have been deleted the PID record should exist, indicate deletion and if possible point to the metadata record.

**PID 17. PID WS**: PIDs need to be used by all parties dealing with data professionally to make full use of advanced opportunities. A PID centric approach to data management, access and use will open the way towards new and comprehensive way of data handling and finally to a Global Digital Object Cloud [5] as a generic, non-proprietary virtualisation layer.

**Handles and DOIs**

**PID10. DOI:** For electronic documents and published digital objects register a digital object identifier (*DOI, which is a Handle with prefix 10*) and associate suitable information with it *(such as citation metadata).*

**PID14. DOI:** A DOI needs to be registered for a published DO and it should be associated with citation metadata.

**PID 22. PID WS**: We urgently need to come to a structured and integrated domain of Handle Service Providers.

**PID 23. PID WS**: Service providers need to ensure that these two interoperable domains are part of one integrated landscape of rich services.

**Others**

**PID 20. PID WS**: Setting up and maintaining trustworthy repositories is key for a structured data landscape guaranteeing access to data and its accompanying metadata.

**PID 21. PID WS**: We need to design the required mechanisms (for facilitating automatic data processing) and build the needed tools now with high urgency.

**PID 24. PID WS**: The PID centric approaches that are key to manage the data Tsunami require simple and clear messages for the users.

Appendix 2



Persistent Identifier (PID) Bundle

1st Draft Version, July 2016

**This document will summarise the GEDE discussions on PIDs, i.e. it will be commented, extended, changed etc. dependent on the group’s progress. Mostly the Wiki space will be used for the interactions. This document will be updated at regular moments.**

**Authors: Peter (RDA), Michael (ENES), Maggie (ICOS), Rob (ELIXIR), Dieter (CLARIN)**

**Specific acknowledgements: Alexander Ntoko (ITU), Larry Lannom (CNRI)**

## **State of Discussion**

During the last decade there is a growing conviction that all digital objects (DO) (data, metadata (of objects), software, queries, ) including abstract objects (concepts, relations, etc.) should be associated with a persistent and globally unique identifier (PID) that can be resolved to useful information such as where to find the associated bit sequences and how to prove identity. In addition to the ongoing scientific debate about the use of PIDs we now can also see an increased interest in industry. It is the Internet of Things with its billions of cyberphysical devices that will create huge amounts of data entities that need to be managed. In this context we can refer to the discussion within the ITU (International Telecommunication Union) on ITU-T X.1255 (which is based on the Digital Object Architecture) and the decision to support the Digital Object Architecture which has as one of its core components a PID system. ITU is focusing on an interoperable platform for the different name spaces that are already in broad use such as Barcodes, RFIDs, OIDs and others.

In the academic and library world PIDs have been discussed and applied to digital data including electronic publications, for over 15 years and several approaches have been made (ARK, Handle, PURL, URN, domain specific solutions, etc.) all having found their specific area of application. Recently published principles, such as those from G8 and FAIR, indicate an acceleration of the discussions and deep recommendations as for example worked out by RDA based on many use cases.

Summarising we can state that

* it is time to work on management and access solutions for digital objects involving industry that are based on standardised PID systems
* it seems that we are moving towards a **Global Digital Object Cloud[[2]](#footnote-2)** which characterizes a virtualisation layer where we primarily deal with stable PIDs and useful metadata descriptions
* there is much experience now in academia and industry for a number of PID systems and their functions in digital object architectures
* there is a general and urgent need for a stable, robust, performant and globally available system that can be used easily by everyone to register and resolve PIDs for different purposes and that can incorporate widely used practices
* there will be specific solutions that have shown their usefulness and it is up to the communities to decide how these practices will evolve and how mappings will be done if necessary

## **Statements about PIDs**

Within the RDA Data Fabric Interest Group (DFIG) we started to collect recommendations related to PIDs that have emerged from the scientific community during the last few years. These can be found under <https://rd-alliance.org/group/data-fabric-ig/wiki/recommendations.html>

Up to now we distinguished a few sources of statements:

1. Suggestions which come from RDA WG outputs (RDA WG)
2. Suggestions which emerge from RDA discussions (RDA)
3. Suggestions from other initiatives such as FAIR, ITU, DOI, etc.
4. Suggestions for RDA Recommendations (RDA REC) which will be the result of RDA interactions

If there are statements that relate to this bundle that come from other initiatives we should of course add them to be comprehensive. The first number is a way to refer to the statement in GEDE discussions, the second specifies the source[[3]](#footnote-3).

**PID1. RDA DFT1.1:** A digital object is ... referenced and identified by a persistent identifier ...

**PID2. RDA DFT1.2:** A persistent identifier is a long-lasting ID represented by a string that uniquely identifies a DO and that is intended to be persistently resolved to meaningful state information about the identified DO.

**PID3. RDA DFT1.3:** A PID record contains a set of attributes stored with a PID describing DO properties.

**PID4. RDA DFT1.4:** A PID resolution system is a globally available infrastructure system that has the capability to resolve a PID into useful, current state information describing the properties of a DO (location, fingerprint, etc.).

**PID5. FAIR-F1**: (meta) data are assigned a globally unique and eternally persistent identifier

**PID6. FAIR-A1**: (meta)data are retrievable by their identifier ...

**PID7. RDA-PIT1**: PID systems should support the generic PIT API where Information Types[[4]](#footnote-4) are openly registered and defined.

**PID8. RDA**: A trustworthy PID system must

* be maintained by a dedicated and reliable team,
* be based on a transparent sustainable business model,
* be provided by a non-profit organisation,
* be subject of regular quality assessments by external parties,
* be governed by international boards,
* be based on open standards,
* be based on a redundant and secure architecture,
* support a huge address space (comparable or even larger than IPv6) and
* support an openly documented API optimally supporting accepted data models.

**PID9. RDA:** The PID Record can be used to store the context of digital objects (bitstream locations, metadata, PID, rights information, landing page, etc.)

**PID10. DOI:** For electronic documents and published digital objects register a Digital Object Identifier[[5]](#footnote-5) and associate suitable information with it (such as citation metadata).

**PID11. RDA**: A PID needs to be requested as early as possible, at least at the time of registration at a trustworthy repository, at which point a PID record needs to be available.

**PID12. RDA:** PIDs are associated with collections which can consist of a number of digital entities, i.e., the level of granularity at which PIDs will be assigned is left to the communities and repositories. A fine level of granularity is recommended to anticipate future applications.

**PID13. RDA:** A metadata description contains the PID of the corresponding object. A metadata description contains the PID of the corresponding object. The PID record can contain one or more corresponding metadata PIDs pointing to different metadata versions, or different components such as for provenance descriptions.

**PID14. DOI:** A DOI needs to be registered for a published DO and it should be associated with citation metadata.

**PID15. RDA**: The PID record should include an expiration date for any digital object which are scheduled to expire. Even for digital objects that have been deleted the PID record should continue to exist, indicate deletion, and if possible point to a metadata record.

**PID16. ITU**: The platform for interoperability of heterogeneous identity management systems is an open architecture based on ITU-T X.1255 and the Digital Object Architecture ... and is capable of to offer interoperability at global level.

**PID17. ITU:** The top 5 benefits from this platform are: 1) framework to enable interconnection of objects, data, devices and processes, 2) in-built security regime (PKI) and data privilege delegation, 3) multilingual support and access to a variety of type value pairs, 4) enable defining new type value pairs for increased flexibility for new types of services and applications, 5) interoperable with existing identity management systems.

## ITU Context

ITU is a global organization, with 193 Member States, 800+ major industry players, academia and other institutions and organizations with the mandate to harmonise and standardise telecommunications and ICTs infrastructure, applications and services worldwide, i.e. scientific usage is just a small fraction of its coverage. ITU acknowleges the advantages of Digital Object Architectures and and its relevance to meet future challenges. In particular the Handle System is seen as a platform that is ready to serve global needs. We can refer to two documents being discussed within ITU in particular:

* Information on the relationships between Recommendations ITU-T X.1255[[6]](#footnote-6) (9/2013), ITU-T X.660 (7/2011)[[7]](#footnote-7) and ITU-T X.672[[8]](#footnote-8) (8/2010)
* Platform for Interoperability of heterogeneous identity management systems (in progress) which analyses whether the Handle system can be used as platform to integrate various identity systems such as 1D and 2D Barcodes and RFIDs

In this note we summarise the ITU findings as follows:

* The Handle System which is a core component of the Digital Object Architecture is compliant to ITU-T X.1255 because this ITU-T Recommendation is based on the Digital Object Architecture and it is governed by the completely independent DONA Foundation based in Switzerland with an international board taking decisions.
* The Handle System has as major characteristics: distributed administration/ownership, enabling interoperability of heterogeneous identity management systems, administration defined per named digital object, secured data binding over public network.
* The Handle architecture allows inclusion of both flat and hierarchical OID structures, can incorporate different OID notations, allows different OID registration authorities to manage the OIDs being created by them, allows owners to associate information with Handles and offers a robust and secure resolution system.
* The Handle architecture can be seen as a platform which enables interoperability of heterogeneous identity management systems such as 1D and 2D Barcodes, RFIDs, GPS etc. on a global scale based on its compliance with ITU-T X.1255 and its embedding in a Digital Object Architecture.

## Application Notes

The range of applications of PIDs in general and in particular Handles and DOIs are commonly known. Some infrastructures such as for example ENES[[9]](#footnote-9), CLARIN[[10]](#footnote-10) and EUDAT[[11]](#footnote-11) based their designs widely on PIDs to cope with the increasing amounts of data and enable stable referencing. Very well-known is the use of DOIs in the area of ePublishing and also publishing data for quite a while. Initiatives such as CrossRef and DataCite offer added value services on top of DOIs. First communities started applying PIDs since about 15 years for their purposes and also research organisations such as the Max Planck Society decided in 2005 already to offer a Handle Service to all its scientists. The publishing industry and the research domain were testing the Handle System, but also other suggestions such as the ARK system for a long period already understanding essential requirements.

In the context of ITU the range of applications of an interoperable identifier platform such as the Handle System is huge and in particular for the IoT scenario a robust and independent system with global dimension will be crucial. Most advanced with industrial applications is probably China where Handles are used to identify digital representations of physical objects in the supply chain of child food to be able to assess state and provenance of each individual can. Millions of PIDs are associated already now giving an impression about the future use of PIDs.

## Conclusions and Recommendations

(this needs to be written after having done the discussion phases)

Appendix 3

Aspects of PID Usage – Discussion Document

Carlo, Maggie, Peter, Tobias, Tibor, Ulrich

30.12.2016

Background

At the GEDE meeting in Bratislava it became obvious that everyone agreed with the crucial role for PIDs in proper data management and access. Yet a number of topics were raised that need to be discussed. It was agreed that these topics should be part of the discussions of the GEDE focus group and that they should be addressed in a session at the next RDA penlary meeting in Barcelona. To move ahead communities want fast answers. Six points were mentioned

1. the related issues of granularity and collection building
	* digital objects (DOs) will be re-used and re-combined by others and we cannot predict how these objects will be used in a few years - this requires to give each scientifically meaningful object an identifier
	* DOs are not just referenced within publications, but increasingly often we will need stable references for our data processing (workflows, etc.) to guarantee reproducibility
	* there will be different strategies dependent on the discipline, the repositories storing data need to make their strategy clear
	* there seems to be a trend that people start assigning Handles at high granularity and DOIs for citable collections (climate modelling, linguistics, etc.)
	* in some labs it is already common practice to create virtual collections which are just some metadata and a whole set of PIDs pointing to DOs; collections themselves get assigned a PID
2. how to do versioning
	* some repositories use an attribute in the PID record to refer to the previous and/or subsequent version; if these attributes are typed also machines can use the information
	* other repositories use metadata records to include this information which is probably not as efficient as using the PID record
3. which kind of metadata to use in PID records
	* there is an urgent need to discuss this - a session should be organised at the Barcelona plenary
	* it is about defining a set of types, but there is no obligation to use them all
	* it is generally agreed that one should not overload the PID record
4. binding role of PIDs
	* it is obvious that we are increasingly dependent on PIDs - thus we need to work towards a stable system that is well maintained, redundant etc.
	* if we have such a system we can use the PIDs to bind various types of information (bit sequences, metadata of different types, landing pages, etc.)
5. semantic categories.
	* there is a need for using Persistent Identifiers for referring to concepts and/or categories used in specific disciplines.
	* it is not obvious which kind of references should be used to refer to semantic categories
	* the semantic web community suggests to use cool URIs
	* there are existing practices in the communities which need to be respected; in biodiversity quite a number of schemes are being used, but yet not in a systematic fashion - they are looking for an overarching schema to overcome fragmentation
6. when to assign PIDs
	* for some digital content it is obvious that they are subject to changes, therefore the question is raised when (small versus major changes) one should assign a new PID to a changed object
	* in some communities people work on such DOs and carry out many changes without “registering” a new version so that it can be accessed etc.
	* possibly the use of versionable databases in conjunction with assigning PIDs to queries - as already suggested by an RDA working group - can address this issue, but not all communities feel this is practical or implementable
	* also in this case the repositories and/or communities need to indicate which policies they follow
	* in some cases it may even be useful to assign PIDs before uploading content into a repository[[12]](#footnote-12) - however then problems may occur (what about relevance and accessibility of data on notebooks etc.)

Discussion

In this section we want to address these six issues and describe answers to provoke further discussions. For some topics several options can be given, however for future machine usage and complexity reduction it is important to narrow down the solution space.

1. the related issues of granularity and collection building

* The large data volumes we are faced with and the type of operations require to work with digital collections. Digital collections are complex Digital Objects, i.e. their content refers to other Digital Objects and they are associated with a persistent identifier and metadata enabling searchability, accessibility and reusability.
* Digital collections can thus be used to combine numbers of other digital objects whose bitsequences can be stored somewhere and to give it a referenceable and citable[[13]](#footnote-13) representation. In this very basic form, collections can help with digital object management even at the beginning of the scientific data life cycle, where the preservation status of objects is still unclear, yet large-scale automated operations must be executed by the data infrastructure machinery. In principle a PhD Student could hide the whole data set that is being used for his thesis behind one Digital Collection and just add one citation to the reference list. However, managing such large and heterogeneous collections may not be a simple task, using stable and coherent sub sets as collection elements in a recursive way seems to be more realistic.
* Using the collection mechanism makes it easy to define granularity at a low level. Since data re-use cannot be predicted it makes sense to define granularity by scientific criteria: a Digital Object should be the atomic set of data values that is scientifically meaningful and that should be referenceable in different contexts such as “a brain scan”, “a video recording of an interview”, “the set of values measured by a sensor in a specific experiment under specific conditions”, etc.
* Fragment identifiers added to PIDs can refer to a subset of data values (such as a number of frames in a video), however they require private resolution operations and cannot be mixed with generally resolvable references and citations. In addition, later operations would allow to create more granular digital objects, if this would be required by science.
* Of course there are problematic use cases such as dynamic databases where cells are filled in and changed at random or sensors generating endless streams of data where segmentation is being introduced arbitrarily. In such cases granularity needs to be defined by applying pragmatic criteria.
* Granularity should be defined by what scientifically meaningful sets of values are that will be used together. Granularity can be adapted at later stages where it turns out to be necessary.
* Important is the availability of a flexible collection builder that allows every researcher to easily combine digital objects (virtually) to sets of data that will be subject of operations or references.
* Communities and or repositories need to make statements about how granularity is being chosen to facilitate re-use.

2. how to do versioning

Versions of Digital Objects are always different Digital Objects, i.e. they have a different checksum, thus are associated with different PIDs and metadata. However different models need to be distinguished.

Model 1 (versioning)

* In this model Versions come into being due to content changes over time.
* Creators may want to express the inherent sequential logic behind versioning by attributes in the PID record. Specific attributes with a well-defined type could refer to the PID of a previous version and another one to the PID of a later version. Humans and machines could make use of this typed information.
* Some communities prefer to store versioning information as part of regular metadata that is also used for search and discovery task. Such an approach may reduce accessibility and persistence of versioning information. A better approach is to store basic versioning information (typed links between objects) close to the objects and PIDs, and keep more sophisticated information (such as associated agents or descriptions of content changes) as part of the more loosely coupled metadata.
* There are many Digital Objects such as a lexicon for a specific language which are inherently dynamic, i.e. the lexicons are continuously being updated. It is a matter of policy when a new version will be registered as an "official release", i.e. many changes will be carried out before a new version will be made public.

Model 2 (representation)

* In this model we assume the existence of a digital work which is represented in different expression formats such as an image encoded as JPG, TIF, PNG, etc. They are different Digital Objects with different checksums, which seems at a first glance similar to versions, but should not be confused.
* The question of representations is orthogonal to versioning aspects and should be modelled as such. For every distinct object version, there can be multiple representation formats, and the choice of formats can also differ between versions (for example, an obsolete format may vanish and be replaced by a new one).
* There are different ways to indicate the relationships with the help of PIDs. One way would be to create a collection with PIDs pointing to the collection members. Here again a type such as “presentation\_type” could allow humans and machines to find the different elements. Type information in the PID records would encode which encoding system is being used to implement for example content negotiation strategies.

3. which kind of metadata to use in PID records

A few initiatives within RDA are now active to define useful metadata categories that could be included in the PID record. GEDE should participate in the coming session at the RDA Barcelona plenary and prepare in a discussion the kind of information GEDE members would like to find in the PID records. A few general recommendations can be made.

* PID records should include so-called “state information” which includes metadata elements that describe the state of a digital object (some people also use the term “system metadata” in this context.
* PID records should include so-called “binding information”. The assumption is that we have persistent PID resolution systems in place. If we rely on robustness and persistence, we can add essential information to the PID such as where to find the bit sequences, where to find the metadata, where to find provenance information, where to find landing pages, where to find rights information etc.
* PID records should not be overloaded to not slow down the resolution machinery[[14]](#footnote-14). However, adding pointers and state/binding information just means to add a few bytes which will not harm functioning. Furthermore the definition and use of more complex types may shift the workload from the PID server to the client side.
* The process should define a number of attributes that can be included in the PID records.
	+ Not every service provider has to use all of the defined types, but if they want to integrate a certain information type they should use the well-defined types to create interoperability.
	+ In addition to the defined types service providers could add additional ones, however, these are not agreed upon and machines will have problems to interpret them.
	+ Obviously we need a body within RDA to decide about which information types should belong to the types registered in a Data Type Registry.
* It seems that there is a trend that service providers are interested in their special profiles (their sets of attributes supported, recommended, excluded and/or required), that could be defined on a prefix level. Such a profile could be made public and transparent as an accordingly defined information type of the prefix record. This should be standardised and the Barcelona plenary could be used to make steps here.

4. binding role of PIDs

As already introduced under 3 an increasing number of repositories is relying on the robustness and persistence of the PID resolution system and it is obvious that we need to do even more to guarantee sustainability. Given that we can rely on robustness and persistence, we can add essential information to the PID such as where to find the bit sequences, where to find the metadata, where to find provenance information, where to find landing pages, where to find rights information etc. Such information is crucial for finding, accessing, re-use etc. Consequently it becomes also crucial to guarantee the stability and interoperability of the Type Registries, where the types that deliver this information are defined.

5. semantic categories

The Semantic Web community was and is highly active and defined a number of mechanisms to efficiently deal with complex semantics. A whole range of web-based knowledge representation standards have been designed and are in use such as RDF, SKOS etc. Technologies to aggregate semantic assertions such as triple stores and to do complex queries on such triple stores such as SPARQL have been developed and are being optimised. Finally, the Linked Open Data needs to be supported.

One of the core principles in these semantic web technologies is that URIs are being used to uniquely and persistently refer to the entities amongst which are semantic categories defined in category registries such as for example Dubin Core (example: <http://purl.org/dc/elements/1.1/contributor>). The definitions of the category "contributor" defined in the DC name space can be found using the specified URI which is using the PURL uniform resource loator to redirect to thebundle location of the requested web resource. In principle PURL is an indirection method as the use of the Handle System is. It shifts persistence from URLs which are known to be not persistent to the PURL service, i.e. if the URL of a web-location is changing it requires just a change in the central PURL database. We need to rely on the persistence of the PURL redirection service. Other communities such as the language community relied on the persistence of the URL for their categories since it is granted by ISO (<http://www.isocat.org/>).

bundle

While the PURL service for example uses the standard redirection method to generate the URL the Handle System requires to specify the location attribute it needs to be resolved to.

As indicated, there are many existing practices for referencing to semantic categories in the different scientific disciplines as well.

This approach is closely related to the metadata categories discussed in section 3. In particular versioning as in chapter 2 can give an example of this relation: a derived version of an object could have a typed entry of 'isNextVersionOf' together with its value. This way would result in a triple of PIDs referring to the new DO, the type definition of 'isNextVersionOf' and the old DO as S-P-O triple.

A discussion at the Barcelona plenary session could be used to discuss this issue in detail. Whatever is being done it needs to be compliant with the Semantic Web suggestions and needs to accept the existing solutions.

6. when to assign PIDs

A number of recommendations can be given:

* Digital Objects should be registered (assigned a PID, store pointers to metadata etc.) once it is uploaded to a trustworthy repository, i.e. it is the repository that requests a PID and updates all related attributes.
* In general it does not make sense that individuals request PIDs, since they cannot guarantee accessibility and stability of data.
* Repositories need to define their policies when and how they assign PIDs and update attributes stored together with PIDs.
* Researchers should be motivated to register their data (and other digital types) as early as possible to enable proper referencing and citing. However in case of versioning of dynamic data an explicit release strategy makes sense. Dependent on the types research communities should specify recommendations.
1. What is meant here that we should not include assertions from individuals, but from groups of experts in this area. [↑](#footnote-ref-1)
2. This term resulted from the discussions on concrete configuration building in RDA’s Data Fabric Group. [↑](#footnote-ref-2)
3. DFT = RDA Data Foundation and Terminology Group (<https://rd-alliance.org/groups/data-foundation-and-terminology-wg.html>) ; PIT = RDA PID Information Types group (<https://rd-alliance.org/groups/pid-information-types-wg.html>); FAIR = <https://www.force11.org/group/fairgroup/fairprinciples> [↑](#footnote-ref-3)
4. Information Types are attributes associated with a PID describing DO properties. [↑](#footnote-ref-4)
5. Technically spoken DOIs are Handles with prefix 10 [↑](#footnote-ref-5)
6. X.1255 is a framework for the discovery of identity management information and it describes an open architecture framework based on the Digital Object Architecture where a PID system is central. [↑](#footnote-ref-6)
7. X.660 defines a tree structure that supports international object identifiers (OIDs) [↑](#footnote-ref-7)
8. X.672 specifies an OID resolution system. [↑](#footnote-ref-8)
9. <https://is.enes.org/> [↑](#footnote-ref-9)
10. <https://www.clarin.eu/> [↑](#footnote-ref-10)
11. <https://www.eudat.eu/> [↑](#footnote-ref-11)
12. It may help to define the term "repository" as something "simple": a "repository" is an entity whose primary tasks are to provide services to access digital object content and essential state information, given an object’s PID, and to enable reliable and trusted data management. [↑](#footnote-ref-12)
13. To ultimately enable citation of collections, additional metadata must be gathered. [↑](#footnote-ref-13)
14. Actually the typing has much less impact on the resolution than on minting process, especially if policies about allowed types are involved [↑](#footnote-ref-14)