Ref. Ares(2020)7201787 - 30/11/2020

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

Call: NFRP-2018

(Nuclear Fission, Fusion and Radiation Protection Research)

Topic: NFRP-2018-11

Type of action: CSA

Project:

“Fair4fusion – open access for fusion data in Europe”

D4.2 Report on Interoperability and Access

WP4

Version	Final
Type	Report
Dissemination level	Public
Work package	WP4 - Open Data Foundations for Open Access of Fusion Data

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

Lead Beneficiary	NCSR-D
Due Date	November 30, 2020
Date of submission	November 30, 2020
Project Name	Fair4Fusion – open access for fusion data in Europe
Grant Agreement	847612
Project Duration	September 1, 2019 – August 31, 2021

Document Information

AUTHOR

Author	Organisation	Contact
Iraklis Angelos Klampanos	NCSR-D	iaklampanos@iit.demokritos.gr
Spyridoula (Iris) Xenaki	NCSR-D	-
Andreas Ikonomopoulos	NCSR-D	anikon@ipta.demokritos.gr
Vangelis Karkaletsis	NCSR-D	vangelis@iit.demokritos.gr
Shaun de Witt	UKAEA	shaun.de-witt@ukaea.uk
Frederic Imbeaux	CEA	Frederic.IMBEAUX@cea.fr

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

DOCUMENT CONTROL

Document version	Date	Author/reviewer – Organisation	Change
1	26/8/2020	I. Xenaki (NCSR-D)	Initial draft of sections 1-5
2	8/9/2020	I. Klampanos (NCSR-D)	Minor edits and updated intro
3	5/11/2020	A. Ikonomopoulos & V. Karkaletsis (NCSR-D)	Major edits throughout
4	10/11/2020	I. Klampanos (NCSR-D)	Updates on PURL and NBN
5	18/11/2020	Shaun de Witt (UKAEA)	Minor updates and included section on interoperability outside the community.
6	19/11/2020	Frédéric Imbeaux (CEA)	Section on interoperability
7	23/11/2020	I. Klampanos (NCSRD)	Drafts of sections 2.7 and 3.1 based on WP4 discussion
8	24/11/2020	Frédéric Imbeaux (CEA)	Review of the document + additional contributions to sections 3-4-5
9	25/11/2020	I. Klampanos & A. Ikonomopoulos (NCSR-D)	Minor edits
10	30/11/2020	I. Klampanos (NCSR-D)	Incorporated suggestions by internal reviewers D. Coster (MPG) and J. Decker (EPFL)

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

DOCUMENT DATA

Keywords	**PIDs, identifiers, FAIR, costing policy**
Point of Contact	Iraklis Klampanos, NCSR-D
Delivery Date	30 November 2020

Table of Contents

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

Executive Summary

Data interoperability that leads to FAIR and Open Data in the fusion community is examined under the double prisms of persistent identifiers (PIDs) and the employment of fusion data, not only across fusion sites but also with respect to the wider fusion community. Besides, the identification of available PID implementations has been supplemented by the cost of globally resolvable PIDs related to publicly available data. The study has recognized that distinct experimental sites maintain data in different formats. Recommendations on PID granularity, PID services and export formats for the general public are provided.

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

1 Introduction

This document reports on the utilization of PIDs – a vital component of FAIR data – in the context of data employment in fusion studies. A number of options are discussed with respect to the visibility and application granularity of these persistent identifiers, while their application costs are presented and discussed. Moreover, recommendations are provided on data interoperability when it comes to external users. Components of a data translation service will be identified, while aspects related to a fusion Open Data platform will be studied in the framework of the Blueprint Architecture (D3.2).

The works performed have evolved in two directions, namely the performance of a survey on persistent identifiers as well as the fusion data interoperability both across fusion sites and in relation to the wider community. The available infrastructures and associated costs for using globally resolvable PIDs related to publicly accessible data have been identified and discussed taking into consideration the data citation needs of the fusion community.

The study has taken under consideration that when it comes to fusion data reusability, the data sets maintained at various sites are not in a standard format, but rather multiple formats may be employed at any one site including proprietary and in-house implementations. It should also be noted that policies concerning the dataset interdependence and upgradability might differ between sites. In addition, a number of recommendations regarding the data format are outlined for the benefit of potential users outside the fusion community and/or those standing at the interface between data analysts and modelers.

1.1 The role of data identifiers in science

Scientific data grows rapidly and relations among these datasets and other resources become increasingly important for furthering science and for improving society. That applies pressure to Research Institutions to develop a strategy for the long-term preservation of their scientific resources to ensure its long-lasting accessibility.

It is therefore increasingly necessary to the scientific community that their resources are registered in well-kept repositories whose content is non-modifiable and may be referenced and cited accordingly. It is also an essential component of the scientific process that older data could be retrieved for comparison or further analysis. While the references made have to be stable, the hosting repositories like any living organism are susceptible to migration paths that include modifications of their hardware, software, physical location or format.

Science has relatively recently started to make heavier use of novel methods to reference scientific data in order to name it in a unique and timeless way similar to the International Standard Book Number (ISBN) for a book that represents a permanent and citable reference to each book identified by it. The resolution of such unique and persistent identifiers mandates a commonly agreed process, while the needed global resolution service ought to have high degrees of robustness and reliability in the long-term due to the importance of the reference resolutions to actual URLs for a high number of transactions.

To this end, several kinds of PID systems are nowadays available and one needs to decide which one of them is most suited to his/her needs. The most frequently used systems are the:

- **Persistent Identifier for the eResearch (ePIC) System**

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

- **Digital Object Identifier (DOI) System**
  - both based on the **Handle System** that is a general-purpose identifier resolution system
- **Persistent Uniform Resource Locators (PURLs)**
  - A URL-based system that points to a resolution service. PURLs have been criticized for being URL based and therefore susceptible to issues arising due to domain name registrars, for instance.

Each of these approaches has a unique set of properties, strengths and weaknesses. In the following section, information and recommendations on the employment of persistent identifiers in digital data that will lead to a globally unique, unambiguous and permanent identification of a digital object are presented and commented.

It should additionally be noted that some disciplines have attempted to make use of their own, in-house, PID service. Most notably, the IVOA (International Virtual Observatory Alliance) went into great detail[1] into the requirements, policies and implementation and these were widely used for some time within the community. However, due to funding restrictions, while support is maintained, IVOA identifiers are no longer regularly minted. However, this is not a recommended approach generally as it makes it difficult to interact with external systems such as CrossRef[2] or Scopus[3].

2 PID Systems and Providers

The most popular PID systems are the core Handle systems, as the ones ultimately provided by ePIC and the Document Object Identifier (DOI) system, e.g., as it's being implemented and offered by DataCite. Functionally, these systems are close, however DOI has a more mature business model on top of the Handle system and it typically offers registration and other added-value services. On the downside, it is more expensive, and its cost depends on the number of registered items. These and other systems are described, in brief, below. Another succinct comparison of the most important PID systems is provided by the *Clarin* research infrastructure[4].

2.1 Persistent Identifier for the eResearch (ePIC) System

[5] A persistent identifier (consists of a prefix and a postfix separated by a forward slash (/) taking the form: <prefix>/<suffix>. The prefix is a single number and groups one, or more, postfixes maintained by an institution or individual. The postfix is a unique alphanumeric string of characters. The combination of the pre- and postfix uniquely identifies a specific piece of data on a specific location by a URL. An example PID is of the following form:

11304/2e873bd8-b988-11e3-8cd7-14feb57d12b9

[1] https://www.ivoa.net/documents/IVOAIdentifiers/20160523/REC-Identifiers-2.0.html
[2] https://www.crossref.org
[3] https://www.scopus.com/home.uri
[4] https://www.clarin.eu/content/comparison-pid-systems
[5] Image from: https://www.pidconsortium.net/

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

With a PID, the actual data is yet to be retrieved. In order to achieve this, a PID resolving service is required, as described in the next sub-sections.

Typically, a PID does not only provide the location URL of the data the PID is referring to but also, in a number of cases, additional metadata is provided as well, e.g., the identifier type, owner, etc.

2.1.1 Resolving PIDs – The Handle System

[6] The EPIC PID service provides unique persistent identifiers and accommodates the management of the metadata accompanying the PID. To actually resolve a given PID to its corresponding location, a PID resolving service is required. For a PID registered with the EPIC PID service, the Handle service run by the Corporation for National Research Initiatives (CNRI)[7] provides these locations based on the PID.

The Handle System is a comprehensive system for assigning, managing and resolving persistent identifiers for digital objects and other resources on the internet, while it includes an:

- open set of protocols
- identifier space
- implementation of the protocols

Below there is a graphical depiction (Figure 1) of a typical PID workflow that employs the Handle Service.

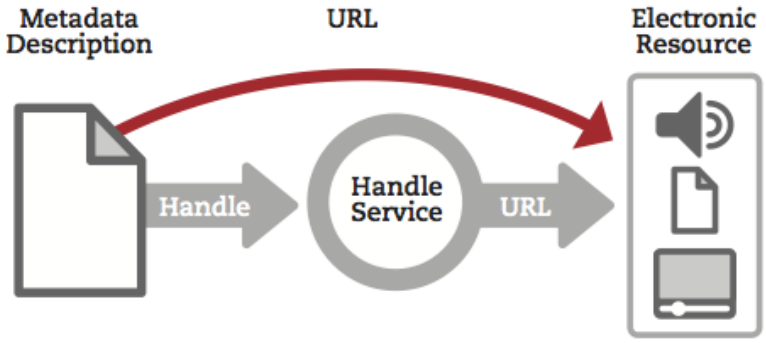

Figure 1: Virtual description of PIDs workflow using the Handle Service (Image from https://www.pidconsortium.net/).

2.1.2 [8]EUDAT Ltd

The EUDAT Collaborative Data Infrastructure (or EUDAT CDI)[9] is one of the largest infrastructures of integrated data services and resources supporting research in Europe. It is sustained by a network of more than twenty European research organizations, data and computing centers that, on September 2016, signed an agreement to maintain the EUDAT CDI for the

[6] Image from: http://handle.net/
[7] https://www.cnri.reston.va.us/
[8] Image from: https://eudat.eu/
[9] https://eudat.eu/

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

next ten years and in 2018 supported the establishment of the limited liability company, EUDAT Ltd. This infrastructure and its services have been developed in close collaboration with more than fifty research communities spanning across various scientific disciplines and involved in all stages of the design process.

In order to access a data object stored in EUDAT, an associated PID is needed. EUDAT is a combined solution of the Handle system and the ePIC service. The B2HANDLE service enables EUDAT services and user communities to assign PIDs to different kinds of managed objects stored in the EUDAT CDI. PIDs are used in EUDAT to identify and cite data objects over a long period of time making it a vital part of long-term data management. Furthermore, data can be directly retrieved by PIDs and corresponding key-metadata can be stored together with them in the, so-called, PID entry.

The B2HANDLE service encompasses:

1. management of identifier namespaces (Handle prefixes)
2. establishment of policies and business workflows
3. operation of Handle servers and technical services
4. a user-friendly Python library for general interaction with Handle servers and EUDAT-specific extensions

The underlying technology of B2HANDLE is based on the Handle System, a reliable, redundant and scalable system built on top of an open architecture. B2HANDLE is mostly transparent to the end-users, especially shielding them from the complexity of infrastructure details. B2HANDLE is a distributed service, with the organisations hosting the service mirroring each other's PIDs. This ensures the sustainability and reliability of PIDs in the EUDAT domain. The mapping process of EUDAT services to the B2HANDLE functionalities is graphically illustrated in Figure 2.

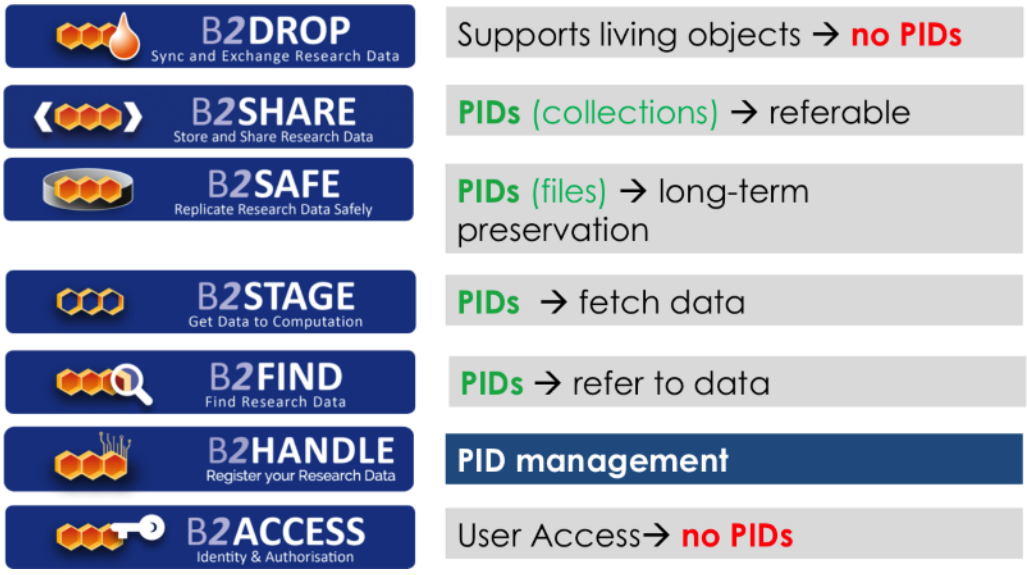

Figure 2: Mapping of EUDAT services to the B2HANDLE functionalities that they each use.[10]

[10] Image from: https://www.eudat.eu/services/userdoc/b2handle

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

In the EUDAT ecosystem, EUDAT services make use of the B2HANDLE service in order to:

- guarantee data access
- long lasting references to data
- facilitate data publishing

On the other hand, the B2SAFE and B2SHARE services use the service to create and manage PIDs for their hosted data objects, whereas the B2FIND and B2STAGE services use the resolving mechanism of B2HANDLE to retrieve and refer to objects.

2.2 Digital Object Identifier (DOI) System

[11] DOI stands for "digital object identifier" meaning a "digital identifier of an object". A DOI name is an identifier (not a location) of an entity on digital networks. It provides a system for persistent and actionable identification and interoperable exchange of managed information on digital networks. A DOI name can be assigned to any entity — physical, digital or abstract — primarily for sharing with an interested user community or managing as intellectual property. The DOI system has been designed for interoperability accommodating usage, or working with, existing identifier and metadata schemes. DOI names may also be expressed as URLs (or URIs) and are of the form <10.xxxx>/<suffix> while they are resolvable using https://dx.doi.org/.

The DOI system was initiated by the International DOI Foundation (a not-for-profit, member-based, organisation initiated by several publishing trade associations) in 1998 which was later standardised as ISO 26324[12]. Users may join a service offered by a DOI Registration Agency[13] by registering material with one of them or developing a community to build one. Existing DOI names can be resolved free of charge while the cost of registering new DOI names depends on the services using a DOI that are provided by a Registration Agency. Each Registration Agency is free to offer its own business model in compliance with the overall DOI policies and individual Registration Agencies adopt appropriate rules for their community and application.

Millions of DOI names have already been assigned through a growing federation of Registration Agencies, world-wide. For example, the

- Crossref[14] application is used by more than 4800 publishers and societies to enable cross-citation of scholarly publications
- DataCite[15] is an international federation of data centres that uses the DOI system
- Entertainment Identifier Registry[16] applies DOI names to film and broadcast assets

The DOI system implements the Handle System along with the indecs Framework[17]; a generic ontology-based contextual data model structure.

[11] Image from: https://www.doi.org/
[12] https://www.iso.org/standard/43506.html
[13] https://www.doi.org/registration_agencies.html
[14] https://www.crossref.org/
[15] https://datacite.org/
[16] http://eidr.org/
[17] https://www.doi.org/factsheets/indecs_factsheet.html

Unique identifiers (names) are essential for information management in any digital environment. Identifiers assigned in one context may be encountered, or even re-used, in another place (or time) without consulting the assigner who cannot guarantee that his assumptions will be known to someone else. Thus, *persistence* of an identifier may be viewed as an extension of this concept: interoperability with the future.

Since the services outside the direct control of the issuing assigner are by definition arbitrary, interoperability implies the requirement of *extensibility*. Hence, the DOI system is designed as a generic framework applicable to any digital object while providing a structured, extensible means of identification, description and resolution. The entity assigned a DOI name can be a representation of any logical entity.

Like the ePIC Handle, DOI's are based on the Handle system and the two can be used interchangeably. Also, the majority of EIROforum[18] members who have adopted the use of persistent identifiers for data have selected the DataCite DOI system.

2.3 URN:NBN

The National Bibliography Number (NBN)[19] is a group of systems employed by national libraries in a number of, predominantly European, countries. The NBN is used to identify content held by these libraries. The Uniform Resource Name (URN)[20] is a scheme for globally unique persistent identifiers and designed to be available and unchanged for long periods of time. Certain libraries have adopted the URN as a means to assign NBNs and provide resolution services. This means that national library records are persistent, globally unique and resolvable. However, this scheme implies ownership of publications and metadata from national libraries, which is not typically the case for scientific data, and it's not globally adopted. A sample URN:NBN identifier (taken from Wikipedia) is `urn:nbn:de:bvb:19-146642`, where the first parts indicate the country (DE: Germany) and the region (BVB: Bavaria).

2.4 PURL

The Persistent Uniform Resource Locator (PURL)[21] is a URL that is used to redirect to the location of a resource on the Web. As the name suggests, the PURL is persistent, while the location of the resource may change over time. The resolution takes place via resolution services. Even though PURLs provide permanence regarding the URL, they are not particularly well suitable for long-term archiving of data. This is due to the fact that PURL resolvers do not know about one another, let along that they offer no guarantees about the resources they point at – they may change, get updated, or even removed without warning.

[18] https://www.eiroforum.org
[19] https://en.wikipedia.org/wiki/National_Bibliography_Number
[20] https://en.wikipedia.org/wiki/Uniform_Resource_Name
[21] https://en.wikipedia.org/wiki/Persistent_uniform_resource_locator

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

2.5 Comparison of PID Systems

PID System	DOI	URN:NBN	ePIC Handle	PURL
Application	Designed for reference and citation of material while they are known and accepted as best practice in research communities	Primarily designed for identification and less so for citation purposes	An all-purpose persistent identification system that is generally useful in assigning identifiers to a large number of digital objects	Designed to maintain permanence of URLs on the Web
Policies	Clear (and strict) policies for their variety of persistent identifiers	Clear (and strict) policies for their variety of persistent identifiers	Quite free allowing users to create their own policies regarding granularity, registration, etc.	No strict policies enforced by default
Repository requirements	There is no requirement for digital objects to be in a Trusted Digital Repository, i.e. one that is guaranteed to remain unchanged, but it is recommended	It is required that objects are housed in such a repository	There is no requirement for digital objects to be in a Trusted Digital Repository but it is recommended	No specific requirements for the resources pointed at by PURLs
Kind of objects to be applied	They are, in general, suited to pointing to data and documents.	The emphasis is on the sustainability of publication, data(sets) and accompanying metadata. Thus, making it ideal to link to publications	It may be applied to virtually any type of object	It may be applied to virtually any type of object
Metadata requirements	It requires mandatory metadata elements	It points to a landing page where the user may publish, or edit, the metadata for an object	It does not require specific metadata	It does not require specific metadata

Content change ability	In principle, it requires the location of objects to remain the same, while it may change the contents of a dataset.	In principle, it requires objects that remain the same, while A new identifier will be needed for the new version of the dataset Points preferably to publications and data(sets)	If a user wants to create different versions of a dataset while using the same Persistent Identifiers, then Handles offers this possibility Designed to work with both collections and individual objects	It makes no assumptions on the underlying resources.
Deleting options	User persistent identifiers cannot be completely deleted and will redirect to a landing page explaining the type of object which was deleted	It does not offer an easy option for deleting persistent identifiers and the service provider must do this for the user.	An administrator can delete the user persistent identifiers.	The owner can delete the content pointed to by a PURL
Resolution	Uses centralised global resolution. It always directs to a landing page with information on the object	Has servers that resolve locally and nationally It always directs to a landing page with information on the object	Uses centralised global resolution. It can point to an arbitrary location, such as a landing page, a digital object or a physical object.	Potentially decentralized solution, with resolvers not knowing one another.

Due to their characteristics outlined above, only Handle-based systems and the corresponding providers are being considered for Fusion data, due to their stability, availability, diversity in the requirements they can satisfy and compatibility with related European platforms and initiatives.

2.6 Monetary and Infrastructural Costs

2.6.1 eResearch (ePIC) System (SURFSara-B2HANDLE)

The cost is independent of the number of PIDs requested. The charges are based on the issuing and maintenance of a prefix instead. This appears to be the most compatible approach with current European efforts, since it is provided as an EUDAT/EOSC service.

The current costs are:

- €42 one-time administration costs
- €42 annual maintenance costs

- €609 annually for PID hosting

2.6.2 DataCite

This is a DOI service that is based on the Handle service for the PIDs[22] where the costs are listed in the Table below for different Tiers.

Tier	Annual DOI range	Organization Fee	DOI Fee	Annual service fee per organization
Tier 1	0 - 1,999	500€	Graded tier 0.80€ per DOI	500€ + 0.80€ per DOI
Tier 2	2,000 - 10,000	500€	1,600€	2,100€
Tier 3	10,001 - 100,000	500€	2,500€	3,000€
Tier 4	100,001 - 250,000	500€	3,500€	4,000€
Tier 5	250,001 - 1,000,000	500€	8,500€	9,000€

2.6.3 EUROfusion-managed ID resolution

A EUROfusion-managed ID resolution could be built on top of either the Handle system or any other, even ad hoc, solution. The proposed implementation would require that, for data – or metadata – hosted on the EUROfusion data portal in the future, a EUROfusion system guarantees the uniqueness of the assigned PIDs to the incoming data. Such a service would be free of charge but it would require implementation and maintenance activities to be centrally managed by EUROfusion.

2.7 Recommendation

Taking into account monetary and infrastructural aspects of the systems above, our recommendation would weigh in favour of using PIDs as they are issued by the ePIC handle system. This would immediately facilitate European interoperability, but also more globally than ePIC PIDs are resolvable also by DataCite, which uses DOIs, and vice-versa.

[22] https://datacite.org/pricelist.html

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

3 PIDs for Fusion Research

3.1 Current Practice

EUROfusion experiments do not use PIDs formally at the moment. Internally they use a more-or-less uniform scheme for identifying experimental datasets based on shot and a version number for processed data (equivalent to the “sequence” number at JET and the “run” number in the ITER Integrated Modelling & Analysis Suite (IMAS)). The resolution is ad hoc and experiment-specific in the sense that there is not a standard way of retrieving data, a standard landing page across EU fusion experiments, etc. The use of standard PIDs will allow EU fusion experiments to reference and exchange data products within and outside the fusion community and integrate with other European and global initiatives. This is, thus, a key improvement towards FAIRness of EU fusion data.

3.2 Recommendation for PID granularity

One of the key issues for assigning PIDs is to agree a granularity at which they should be supplied. Across fusion, most APIs relate to shot/pass (where shot is a single plasma discharge of an experimental device and pass is effectively a dataset version number) and thus, this would seem to be the clear choice. However, many if not most analysis is not done across a whole shot but may only use a subset of the diagnostic information available and even then only specific time-slices from each shot. There are certain concepts, such as the ‘flat top’ where the plasma is in a pseudo steady state which is often used for analysis, but the definition of this flat top period is not standardized and often left up to the researcher or team. Once ITER data becomes available this issue will be exacerbated as it will likely have many ‘flat tops’ due to changes in the tokamak parameters during an experimental run.

In addition, in many devices there are specific experimental campaigns which generate several shots, each of which may have many versions, but shots can be performed outside of any specific campaign for many purposes. This is shown Figure 3, below. Other researchers are interested in the ramp up phase or looking for specific items such as disruptions which may only appear in some signals.

This project has received funding from the European Community's Horizon 2020
Framework Programme under grant agreement 847612

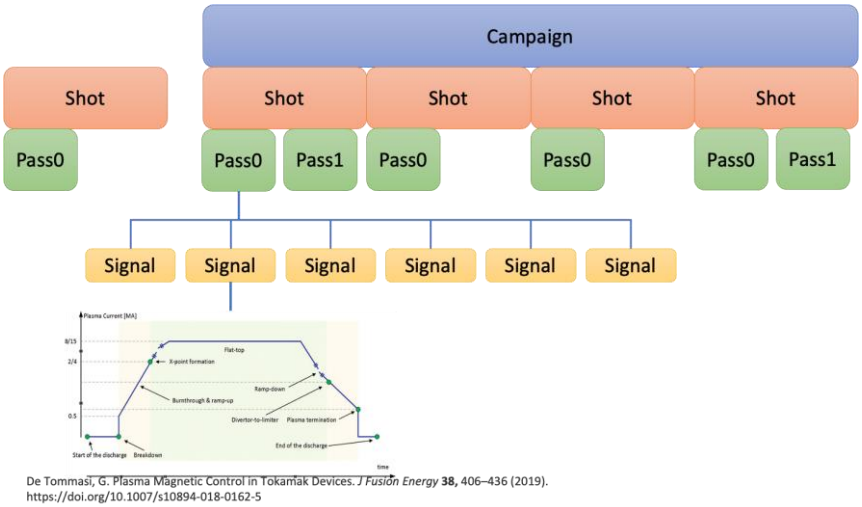

Figure 3: Possible PID Granularities

For the purposes of this project we have initially agreed that the ‘pass’ or ‘run’ number is the appropriate granularity as it represents the current top level API granularity. If this is made common, then each device would also be able to apply identifiers at different levels dependent on their user base and specific use cases.

In addition to the “shot-based” data, fusion experiments produce also “continuous data”, related to technical systems that function and acquire data both during and between plasma operation. Typical examples of these are barometry, calorimetry and mass spectroscopy which must be monitored on longer time scales than just the plasma discharge. A typical method today is to refer to this data by absolute time instead of “shot” number and the typical granularity is to have datasets covering one day of tokamak operation. This current practice seems also relevant for future machines, thus for continuous datasets we would recommend using the same day-based granularity for PIDs. The version number “pass/run” is also obviously relevant for continuous datasets.

4 Recommendations for Data Interoperability outside of the Fusion Community

Regarding interoperability with formats used by the general public, the IMAS Data Dictionary used within the fusion community will still be the reference for the definition of metadata and data, but it is not easy for a non-fusion researcher to invest time in learning how to use directly the IMAS Access Layer to retrieve data or metadata. The recommendation is thus, to use simple formats such as ASCII to provide at least the metadata to the general public. This can be implemented as the main output method to send back to the user sets of metadata from the portal (e.g. resulting from a metadata query). Beyond metadata, in principle text-based formats (e.g. ASCII) can also be used to dump parts of the physics dataset referred to by the metadata. The IMAS Access Layer has built-in methods to do so, although text-based representations will be quite voluminous

when large data items are requested. The IMAS Access Layer also has an HDF5[23] backend, but the internal structure of the generated HDF5 file remains complex to interpret from the perspective of someone using standard HDF5 tools, without using the IMAS tools. It means that a user inspecting the HDF5 file will not immediately recognize the usual view of the data structure documented in the Data Dictionary. Therefore, using ASCII-based formats, e.g. custom formats or ones based on CSV, JSON and others, and possibly generated on the fly by the IMAS Access Layer, still appears as the easiest way for the general public to access fusion data without having to learn and use specialized data formats and access methods.

Within this project we have attempted to improve the interoperability of the metadata by adding some key elements of Dublin Core to the IMAS Data Dictionary – the community 'standard' for storing and retrieving data and metadata, including scope for a persistent identifier and version resolution. Therefore, once this information is populated at the Fair4Fusion portal level, we can use a metadata harvester to make the metadata accessible through an external portal such as that provided by the European Open Science Cloud.

In addition, we are investigating the possibility of making some of the plots available in the Summary Interface Data Structure (IDS) available directly from the Fair4Fusion portal in a commonly-used text-based format, such as CSV.

5 Conclusions

This report discussed the applicability of persistent identifiers for open fusion data, taking under consideration the WP2 outcome regarding experimental as well as non-experimental fusion data and the policies for accessing it. A number of PID systems were identified, and their delineation revealed infrastructural and implementation requirements for adopting them. In addition, the cost associated with the adoption of PID services has been recognized and listed. To facilitate interoperability within Europe and globally, and taking into account infrastructural and monetary aspects, our recommendation is to consider using PIDs as they are issued by the ePIC handle system. Furthermore, data interoperability outside the Fusion community has been discussed and text-based (e.g. ASCII-based) formats have been identified as a relevant solution in order to ease the data retrieval for the general public, removing the burden of having to install and learn specialized access methods and data formats. Moreover, the IMAS Access Layer as well as the Demonstrators built within this project already have the capability to export metadata and data to ASCII.

[23] Hierarchical Data Format version 5