PIDs as IRO Infrastructure - Early Findings

Rachael Kotarski (British Library; 0000-0001-6843-7960)
Jack Kirby (Science Museum Group; 0000-0003-0334-1954)
Frances Madden (British Library; 0000-0002-5432-6116)
Lorna Mitchell (Royal Botanic Garden Edinburgh; 0000-0002-3977-464X)
Joseph Padfield (National Gallery, London; 0000-0002-2572-6428)
Roderic Page (University of Glasgow; 0000-0002-7101-9767)
Richard Palmer (Victoria and Albert Museum; 0000-0002-3177-1313)
Matt Woodburn (Natural History Museum, London; 0000-0001-6496-1423).

Report Contents

Executive summary

The Persistent Identifiers as IRO Infrastructure project has delivered an initial set of investigations to understand how to take forward persistent identifiers (PIDs) as a core component of infrastructure in support of joining heritage collections together as a ‘National Collection’. Based on these early findings, we have formulated an initial set of 6 recommendations for the next 12 months:

- Recommendation 1: The value proposition of PIDs has been articulated, but needs to be addressed directly to decision makers within cultural heritage institutions. The PIDs as IRO Infrastructure project will create materials specifically for decision makers.
- Recommendation 2: In support of choosing appropriate identifiers:

    - A: the project will develop a description of broad institutional requirements, defined to a set of 4-5 levels of complexity and matched up to the features of various identifiers. This advice will build on the very early definitions contained in this report's List of identifiers.
    - B: Subsequently, the project will create guidance which will outline practical steps to help organisations move between these levels and work with PIDs that support more complex use cases.
- Recommendation 3: We strongly recommend that heritage organisations start to work with their system suppliers to ensure systems meet their PID-based requirements (as defined in Recommendation 2A).
- Recommendation 4: The project and IROs to continue to gather cost information on PID implementation, in particular across a more diverse sample of organisations.
- Recommendation 5: This project to offer some additional guidance to staff working with collections on how citation practices for heritage artefacts could be enhanced with the use of identifiers.
- Recommendation 6: Sector-wide governance and policies for PIDs should be investigated as an option to encourage uptake and to have a coherent approach to implementations and use of PIDs. The PIDs as IRO Infrastructure project will begin this work and make further recommendations on a sector-wide approach, but these will need to be tested and refined by Towards a National Collection's future Discovery Projects.

These recommendations partly give us further direction for the remainder of the project (Recommendations 1, 2, 4 and 5), but some require further work that cannot be achieved within the timeframe or scope of the project. To further this work beyond the project, particularly in support of Recommendations 3, 4 and 6, we propose that:

- Recommendation 7: More IROs, higher education institutions and heritage organisations should implement policies on use of PIDs to support linking of items and their metadata across institutional boundaries, and identify a minimum technical passive provision for PIDs that future-proofs new tools and systems for their use.
- Recommendation 8: Where key strategic systems cannot be easily reworked for PID use:
    - A: the recommendation is for lightweight add on software which can be integrated alongside existing systems.
    - B: This may still be beyond the reach of smaller organisations with little or no technical capacity, and so shared infrastructure approaches in support of such organisations should be explored.
- Recommendation 9: The Towards a National Collection programme, in collaboration with IROs, heritage organisations, higher education institutions and future projects, should explore a sector-wide approach to identifiers that enables the benefits of wider take up, while respecting the diverse internal management needs and processes of organisations.

The Persistent Identifiers as IRO Infrastructure project was launched in January 2020, funded under the UKRI Arts and Humanities Research Council's Towards a National Collection programme. Our aim is to explore persistent identifiers as a foundational infrastructure for the

programme, using their power to provide a long-lasting click-able link to a digital object. They are recognised by UKRI as a tool for enabling data discovery, access and citation[1]. Supporting wider use of PIDs for collection objects, environments, specimens and related items will allow long-term, unambiguous linking of collections that will create a digital National Collection.

This project is bringing together best practices in the use of PIDs, building on existing work and projects. We will share expertise and provide recommendations on the approach to PIDs for colleagues in institutions across the UK heritage sector. Through a mixture of workshops, desk research and case studies, the project will answer questions such as 'What are the gaps in the existing PID landscape for heritage collections, buildings and environments?' and 'What should a PID infrastructure, strategy and governance framework look like for a unified national collection?'.

This report outlines those activities to date and provides an overview of the awareness of PIDs across the sector based on the findings at this stage in the project. It also showcases the requirements of the sector and barriers to adoption we face. In turn, it aims to offer initial solutions to the challenges identified in our activities so far. This report's timing is designed to ensure that Discovery Projects have an opportunity to address some of the questions and issues identified so far.

Early results are reported from:

- **A survey of current PID awareness and use in the UK heritage sector.** Survey results suggest that there is awareness of PIDs as a technology, but the benefits of them to decision makers within the sector are currently unclear. Some more nuanced views also emerged around a lack of PID creation and management within the technical solutions which are in use by organisations.
- **Two initial webinars**. Comments and discussion within the webinars echoed findings of the survey, but underlined the need to articulate the specific benefits of global uniqueness, resolvability and persistence for identifiers.
- **A demonstrator tool.** The Annotate It! tool shows how PIDs can be used to understand where collections connect to research. The significance of this work is in showing that as long as PIDs are in place, the institutions managing collections need do very little to be able to gather information on the links between, and impact of, their collections within research.
- **Two initial case studies.** The key findings from the case studies emphasise an approach to implementing PIDs within lightweight middle-ware, which will allow uptake alongside critical strategic infrastructure.
- **A literature review.** Much of the literature focuses on the benefits of PIDs to the sciences and to research communication. While there are fewer examples of PID use and research at heritage organisations, they do start to identify the specific benefits of PID use to the sector.

[1] As Research Council UK's *Common Principles on Data Policy* (2011), under extended guidance for Principles 3, 4 and 6; and the *Concordat on Open Research Data* (2016), Principle 8.

Background

How can you have a national collection if you do not know what is in it? The Towards a National Collection programme aims to unify collections and this project explores how best to utilise persistent identifiers (PIDs) as a key component of the infrastructure on which the national collection can be built. Heritage collections around the UK are effectively siloed on different websites, with some aggregation services relying on spreadsheets to keep information up to date, a process that can be complicated and time-consuming. Institutions can have varying rules, licences and policies around their collections and the extent to which information about collections is available online can vary considerably.

By providing a way to create a stable and trusted link to objects on the web, persistent identifiers can be viewed as a cornerstone of the infrastructure required to unify collections. They can be used within metadata to connect objects across silos, helping our understanding of the nature and relationships between these collections. As long-term custodians, many independent research organisations[2] have a responsibility to 'persistence', and so that particular aspect of these identifiers presents an opportunity to ensure that whatever solutions are proposed by this project are sustainable.

Persistent identifiers have gained traction in several spheres such as research (Meadows et al., 2019) and the entertainment industry (Kroon, 2014). There is a mature suite of infrastructure which cultural heritage organisations can leverage and enhance to their own purposes. By making the most of existing PID infrastructure, we can embed heritage collections in the wider scholarly digital ecosystem, and take advantage of the metrics and tools that already exist to support the management of those collections.

The longer term aims of the project are to provide recommendations to assist institutions in making best use of the existing infrastructure so that the sector as a whole can realise the benefits of PID use. Many heritage collections are already adopting linked open data approaches that allow collections to be queried and explored in a standardised machine readable way. This can be complemented and enabled by PIDs and their implementation.

In the nearer term, the project aims to understand the state of the art and to try to formulate a position to take persistent identifiers forward within the UK heritage sector. For the Towards a National Collection programme to be a success, some common approaches will need to be adopted by the partners involved. While this project does not aim to be prescriptive, we can produce some useful common criteria which heritage collections can use to enable integration into the National Collection.

Approach

The project has used several methods to assess knowledge about persistent identifiers within the sector and establish the ‘state of the art(s)’, including a literature review, two case studies based on project partners’ use of PIDs, a survey of the sector and two webinars (held in place of workshops).

[2] http://ahrc-iroc.org/

Survey

The PIDs as IRO Infrastructure Survey was launched on 28 May 2020 and remained open until 14 September 2020. The long response timeframe was deliberate, to accommodate responses from any staff who may have been on furlough due to the COVID-19 pandemic. The survey was designed with a series of questions which can be used for benchmarking the experience and familiarity within the sector. It is planned to issue the survey again in Summer 2021 to enable a comparison over time and allow us to see what impact the project has had. The final results will be included in the final report for the project.

In order to try and best capture the state of the art across as broad a swathe of the community as possible, we allowed the opportunity to provide responses as both an organisational or an individual. Responses were invited internationally but it was clear in all messaging that it was primarily aimed at a UK audience. The survey was publicised via the HeritagePIDs Twitter account[3], the second project webinar, the PID Forum, the Towards a National Collection Twitter account and the Towards a National Collection newsletter.

Preliminary findings were shared at a webinar on 17 July 2020, by which point the survey had received 52 responses. The survey was promoted again at this webinar, in follow up mailings and via social media until it closed on 14 September and received 66 responses in total. The survey responses were anonymous but respondents were asked to confirm on behalf of which organisation they were responding. The organisations or affiliations are not communicated in the survey results but general groups are mentioned, e.g. independent research organisations and higher education institutions.

Webinars

Due to the COVID-19 pandemic, in lieu of in-person workshops, the project held two webinars. The first was a launch webinar in April 2020 and a second in July 2020. The launch webinar provided an overview of the project, an example of a use of PIDs from the National Gallery and a potential application from the Heritage Connector project. The second webinar provided an overview of the interim survey results and a demonstrator developed by Professor Roderic Page from University of Glasgow. A panel discussion followed in response to the interim survey results. Panel members included Rebecca Bailey, Programme Director, Towards a National Collection; Lorna Mitchell, Head of Library, Archives & Publications Royal Botanic Garden Edinburgh; and Claudia Fabian, co-founder of the International Standard Manuscript Identifier.

Case studies

Case studies have been identified as being core project outputs from the outset of the proposal as they provide real-world examples of how PIDs can be implemented. To date, two case studies have been completed and are due to be published imminently, one based on the British Library's implementation of PIDs and the second based on the National Gallery. In 2021, the project will also produce case studies on the work of the Natural History Museum in March and Royal Botanic Gardens Edinburgh in June. In addition, smaller case studies will be published as blog posts on organisations and projects which do not yet have implementations

[3] https://twitter.com/heritagepids

of PIDs but have specific requirements around them e.g. Victoria and Albert Museum and the UKRI AHRC funded Linked Conservation Data project[4].

Literature Review

A high level literature review focusing on the benefits of persistent identifiers, especially those related to cultural heritage collections was conducted. While there is a broad corpus of material around the subject of linked data and cultural heritage, the number of articles which discuss persistent identifiers specifically is low.

The state of the art(s)

Survey Findings

Due to the low sample numbers, it is difficult to report strong statistical differences in responses across survey questions, but they do provide us with a quantitative assessment that was previously unavailable. The majority of the survey results are summarised in Figure 1 on page 9.

Survey data is available from the British Library Institutional Repository with identifying responses having been removed, see https://doi.org/10.23636/1210. Due to some geographic locations having only one or two responses, all have been assigned either to 'UK' or 'Rest of the World' to ensure anonymity, but a summary of the locations of respondents is given here.

Respondents

Of the 66 responses 29 were submitted on behalf of an organisation and 37 were from individuals. 57 were received from UK organisations, 46 in England, 7 in Scotland, 3 in Wales and 1 from the Channel Islands.[5] No responses were received from Northern Ireland.

Respondents were asked to describe with which type of catalogue data they work. It was possible to select multiple options with many of those who selected the Other field indicating they were working with Art or Gallery collections. While there are responses from staff working in all major areas of GLAM, there does seem to have been a large number of responses from individuals working with archive collections (37 responses) compared with other collections.

The number of organisational responses received was also somewhat low. This could be explained by the furlough scheme meaning that remaining staff were unable to give a response where information from colleagues would have been required. While many of the organisational responses were from independent research organisations, the number of local and regional institutions represented was low. There could be many reasons for this, including furlough dis-proportionately affecting smaller organisations; our not being able to push the survey through appropriate communication channels, or simply because the level of awareness of PIDs by staff within that sector may be lower than national level organisations.

[4] https://www.ligatus.org.uk/lcd/

[5] International responses are excluded from the survey findings beyond the questions clarifying respondent's backgrounds.

Despite these limitations, we have been able to collect far more qualitative information on the use of PIDs across the UK heritage sector than was previously available, and we reflect on results as quantitative evidence for the way forward.

The types of roles of those responding were quite broad, which is not unexpected given the differences in scale of the affiliation of those responding, however the number of people working in IT within those organisations is high proportion of respondents which indicates that not all staff working in organisations may have an awareness of PIDs.

Awareness

Our starting assumption was that awareness and use of PIDs amongst colleagues in the heritage sector is low. Survey results showed that the awareness of PIDs amongst respondents seemed surprisingly high (23 of 57 responses), as did the number of organisations who said they were actively using PIDs (19 of 27 organisational responses). Given the large numbers of local and regional organisations who are assumed not to have PIDs in use within their collections, there is concern that there is a large swathe of the community that may have missed, or intentionally not responded to this survey and therefore their views are missing. When the survey is recirculated next year, we will aim to attract a broader array of respondents and use broader networks to do so. These results can also be explained by the fact that the survey did not restrict itself to PIDs for collection items but also included utilising DOIs in publishing or encouraging staff to have an ORCID. While these are activities where staff interact with these PIDs, they are not activities where the PIDs are directly related to management of collections.

There was a concern from the outset that the topic of the survey would see the respondents self-select from a group with existing awareness and concerns around PIDs. We actively sought ‘null’ or negative responses, but it is unclear whether this was successful, or whether the awareness responses are representative.

The spread of awareness of the different PID Types (see Table 1: A table illustrating the awareness responses to a range of PID Types. Definitions of the PID types mentioned in the table are included in the List of Identifiers.) was not surprising with disciplinary specific and emerging PID types (e.g. IGSN and ROR) having lower awareness levels than more established PID types (such as DOIs and ORCIDs). The relatively high awareness of URN/URI and PURL speak to the number of local implementations of PIDs which appear to have been implemented or are in the course of implementation by the survey respondents and the fact that those are common PID implementations within the heritage sector. While these implementations are to be commended, they should be used with caution. These identifiers are not necessarily globally resolvable and their persistence may not be guaranteed beyond the project creating them. We have created a list of identifiers that has arisen through the work so far, and indicated to what extent they meet the project definition of identifiers that are globally unique, actionable, and where the global uniqueness and action-ability are guaranteed for the long term.

The fact that the most common response for the interest in PIDs was 'exploring options for new services' implies that PIDs are under consideration in a majority of respondents' organisations, something which is encouraging to the project.

Awareness of PID Types	ARKs	DOI	Handle	ISNI	LSID	IGSN	ORCID	ROR	PURL	URN/ URI
I haven't heard of this	27	6	19	11	42	38	12	38	7	10
I've heard of it, I don’t know what it is	11	7	13	10	2	6	5	4	12	7
I've heard of and seen it in use	13	22	17	26	8	7	16	6	20	16
I use these identifiers	2	22	5	6	3	2	23	4	17	21

Table 1: A table illustrating the awareness responses to a range of PID Types. Definitions of the PID types mentioned in the table are included in the List of Identifiers.

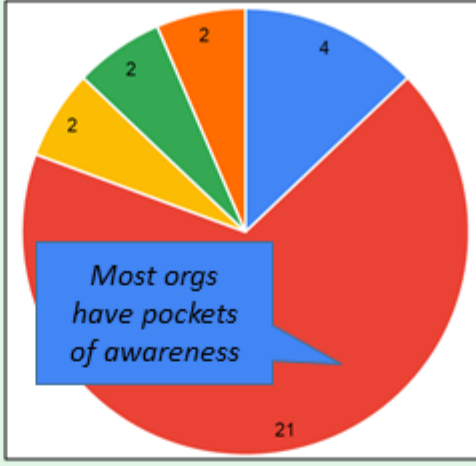
Figure 1 (next page): A graphical summary of the survey results.

Awareness of PIDs

Personal

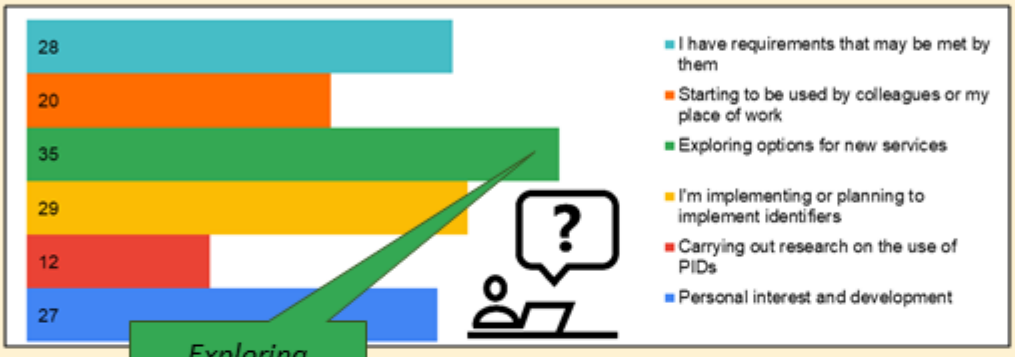
- I have no real knowledge and I haven't used them actively
- I'm aware of a few types but have never used them actively
- I'm aware of and have used some persistent identifiers
- I regularly use them

Organisational

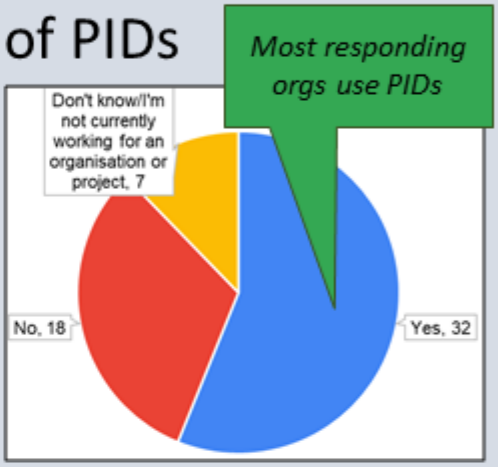


- Low or no awareness
- Some pockets of awareness
- All staff who would benefit are aware
- Broad awareness across the organisation
- Other

Interest in PIDs

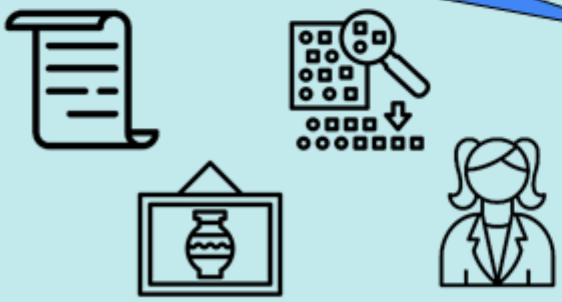


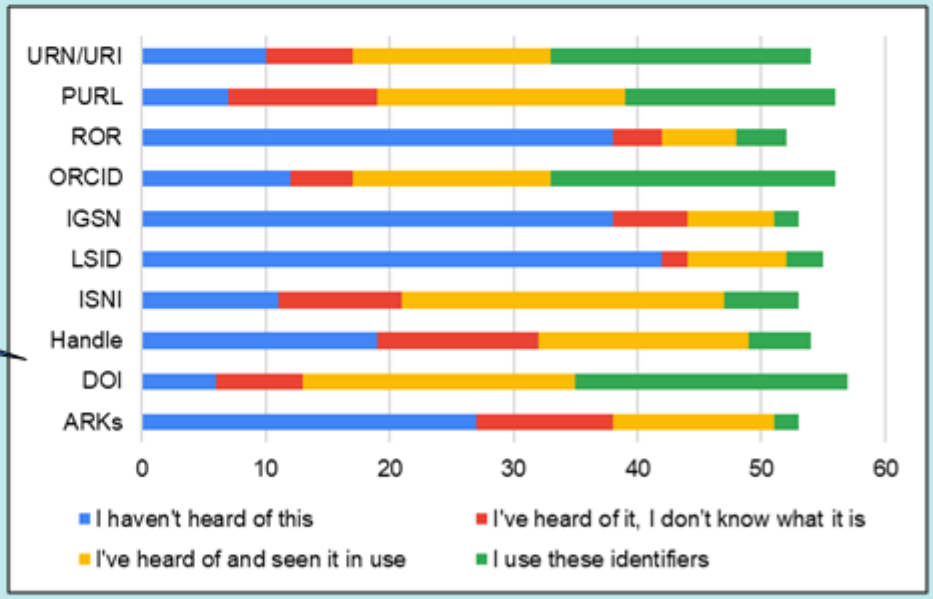
Organisational use of PIDs



Awareness of PID Types

Awareness of mature and generic identifiers was higher than emerging ones





Barriers to PID Adoption

Perceived to be:
Awareness
Resources
Technical

Respondents were also provided with the project's definition of a persistent identifier throughout.

> *"As a reminder, by 'persistent identifier' we mean an ID that is globally unique, actionable (it can be clicked to take a user to a resource or information about a resource), and managed to remain unique and actionable for the long term."*

This led to some responses indicating uncertainty around the definition of PIDs:

> *"Our archives catalogue references are supposed to be persistent but in practice they are not always."*

> *"I'm actually now questioning if I do know what a PID is."*

Organisational Use

Expectation was that organisational use of PIDs is low, but again, we were surprised by the level of organisational use of PIDs reported in the survey. Where users answered 'Yes' to the question of their institution using PIDs, respondents were also asked to provide details of the identifiers used within their organisation. They gave a variety of responses; some of these matched our definition of PIDs above, but some responses included identifiers which would not necessarily match that definition. These responses were analysed and coded based on three categories identified: 'Using global PIDs' (16), e.g DOIs, 'Using a local PID', e.g. URIs (10) and 'not using PIDs' (4). Those using global PIDs most often specified DOIs (11) and ORCIDs (11). This highlights a need to define institutional requirements matched against the features of a range of PIDs, not solely those that match the project definition of globally unique, resolvable and persistent identifiers.

Benefits and Barriers

Respondents were asked to describe the benefits of PIDs in a free text response. When these were analysed three categories emerged: 'Efficiency', 'Trust' and 'Interoperability' and all responses were coded against these. Examples of responses within each category are given below.

> *"Consistent and persistent IDs that can be used across multiple systems - long term efficiency and trust"* - Example of a response highlighting Trust and Efficiency as benefits of PIDs

> *"Promotes easy access, citation tracking"* - Example of a response highlighting Efficiency as a benefits of PIDs

> *"Provide potential for linking data to allow people and services to exploit the connectivity and richness of multiple sources of information, in particular through linked open data."* - Example of a response highlighting Interoperability as a benefit of PIDs

These responses highlight that in articulating benefits to decision makers, we should particularly focus on these three elements.

Respondents were asked to classify the barriers to PID adoption. The barriers to implementing PIDs were in some ways unsurprising with the emphasis on the lack of awareness (44 responses) but Resources (42) and Technical (39) barriers attracted nearly as many responses.

Respondents were asked to further comment on barriers. Based on analysis the following codes were identified and assigned to the responses:

- 'lack of awareness'
- 'doubt of the long term value proposition': as distinct to a lack of awareness of the value
- 'software support': referring to a lack of support in commercial software 'resources': both human and financial.

The free text responses did hint at some other issues including a concern about the value proposition of PIDs, indicating that in some areas at least, while the benefits of PIDs may have been communicated they were still not deemed worthwhile amongst other activities.

> *"Our digital presence is not seen by...senior management as having a research value or a long term value*"

> *"As there is a noticeable cost to start registering PIDs it is harder to start to practically demonstrate how they could be used within an organisation. Also PIDs are generally needed to start to do new things, connect to other organisations, to secure the digital identities of ones collection, as the basis of more complex cross-department or cross-institutional documentation or research resources building. For institutions who are not already trying to do these, it will be harder to argue why they need to change their existing processes which still work. A clear set of procedures that would allow people to slots PIDs into existing systems with minimal development would be ideal. Also having a clear idea as to how the costs might be shared may also help, some bigger institutions may need their own namespace but that will not be true for all."*

The fact that several responses also mentioned the need for tools which support and integrate PIDs is notable and point to a need to engage with commercial vendors as has been done in the DDHN project, discussed further in Literature Review. It was clear from the respondents that the majority of them understood fully the benefits of PIDs, while some definitely approached it with a linked open data perspective. However 26 respondents left this field blank which may mean that their perceived benefit of PIDs was low or they did not have an immediate use case which PIDs would address. Equally this could point to a lack of understanding of the use and benefits of PIDs generally.

Further responses hint at the need for wider adoption with organisations as well as across the sector to realise the full benefits of PID use. If interoperability and linking between

organisational silos is the benefit, these links can only be made to other systems and organisations that have adopted PIDs:

> *"Use has begun, but PIDs are only obviously useful within an institution once there use reaches a certain scale. Before this critical point they are often seen as a costly extra which might be useful later - also technology covering there use might be outside of the current IT scope."*

> *"Persistent identifiers also enable resources in our collections to be cited and reliably linked to, and connected to other resources, held by other institutions, or more generally linked to across the web"*

The final question related to the utility of the proposed outputs of the project. All of the proposed outputs were felt to be very useful with Recommendations on selecting PIDs (41), Recommendations for moving the sector forward (36) and a Toolkit supporting decision making (35) selected 'very useful' by the highest numbers.

Annotate It! Demonstrator

To help make the case for the value of having PIDs for collection objects, we created a live demo for a small set of objects in natural history collections at The Natural History Museum, The Royal Botanic Gardens Kew, and Plantentuin Meise, each of which has a PID in the form of a URL. By adding a Javascript bookmarklet to their web browser, a user can visit the web pages for these objects and discover publications that cite these collection objects (Figure 2). They can also visit the web pages for those publications and discover the reverse links to those collection objects.

The significance of this demonstrator is that neither the institutions managing those collections, nor the academic publishers hosting the articles need do anything to their existing web sites in order for the links to be discovered, the links themselves are stored in a separate database. So long as the collection objects have PIDs, and those PIDs are cited elsewhere (or if not the PID itself, something such as a catalogue number that could be mapped to a PID) then we can demonstrate links between collection objects and what the academic community is saying about those objects.

The missing piece of that puzzle will be the requirements from publishers for citations to the source objects that use PIDs, in the same way they require the same for data citation. However, these are starting to emerge, as seen from the journal Nature (Groom et al., 2017) and Wikipedia[6] also strongly recommends PID-based citations that would support this approach. This was also reflected in the survey:

> *"conservation & heritage scientists nearly always cite DOIs and conservators nearly always don't"*

[6] https://en.wikipedia.org/wiki/Wikipedia:Citing_sources#Preventing_and_repairing_dead_links

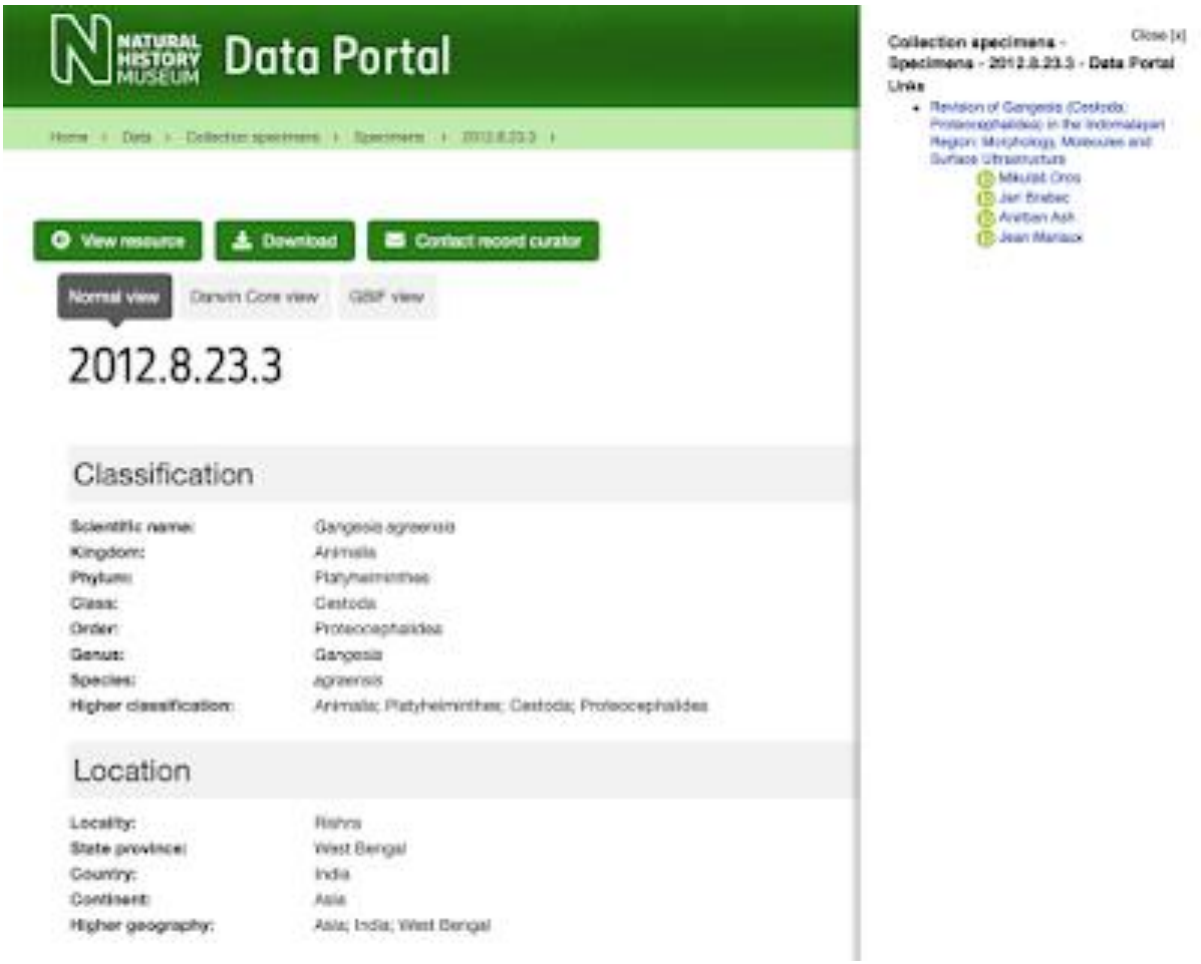


Figure 2: Screenshot from the PID demo showing a specimen in the Natural History Museum (catalogue number 2012.8.23.3, PID https://data.nhm.ac.uk/object/6e8be646-486e-4193-ac46-e13e23c5daef) linked to a publication that made use of that specimen (PID https://doi.org/10.1371/journal.pone.0046421).

A video showing the demonstrator in action is available as part of the materials for the second project webinar, available at https://doi.org/10.23636/1189. Code for the PID demonstrator is available from https://github.com/rdmpage/pid-demonstrator.

Webinar Findings

Both webinars were well attended and each received 187 and 136 registrations and 123 and 94 attendees respectively. An online feedback gathering tool, Mentimeter,[7] was used during both sessions. The responses to the first webinar's feedback are included in Figure 3.

The findings of the webinar are in line with the results indicated in the survey to an extent. There is some awareness of persistent identifiers, but implementations are for the most part, not mature, and the observed barriers to adoption are similar. The main discrepancy seems

[7] https://www.mentimeter.com/

to sit with the difference in the perceived maturity of PID implementations and their use across organisations. This could be attributed to different cohorts responding in each case. In addition the phrasing of the questions may have caused more users to identify as using PIDs in the survey where many respondents mentioned support for PIDs which were not being used for collection items.

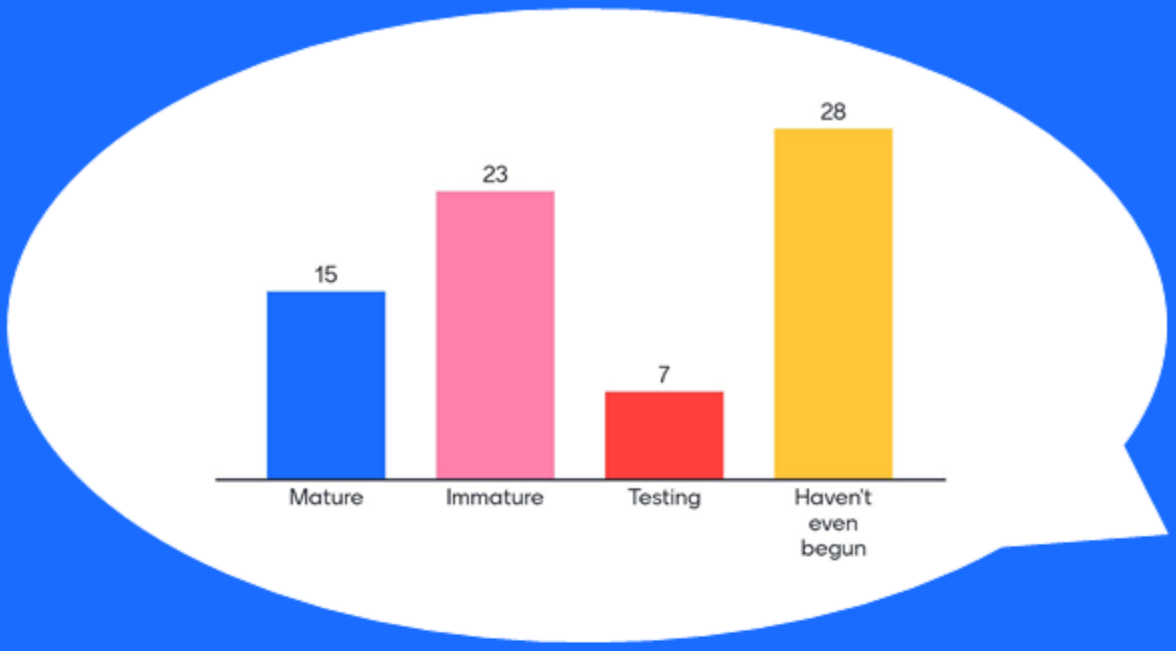
There was a discussion in both webinars around the narrow limits of our definition of persistent identifiers. This observation is supported in responses to the survey, where there was a clear appetite to include identifiers that are not necessarily globally resolvable. The additional benefits of extending PID requirements to fit the more strict definition are not necessarily clear to all, and so there is a need to articulate this against the existing collection management needs and experiences of heritage organisations.

> *"...it will be harder to argue why they need to change their existing processes which still work"*

Figure 3 (next page): A summary of results for questions posed at the launch webinar, held virtually on GoToWebinar on 6 April 2020.

Launch Webinar Feedback

What would you say is the state of PID use at your organisation?



Do you use any PIDs at your organisation?

What PIDs do you or will you use?

In response to the need for expressing benefits, we scheduled such a demonstration for the second webinar. The Annotate It! tool was demonstrated, showing the possibilities for associating publications and their identifiers to related collection items. While the majority of Mentimeter respondents felt that this was a useful way to show the potential of PIDs (50 “Yes” and 15 “Maybe”), the response to an open question of how this can be scaled up drew less positive responses. These included:

> *“We still need PIDs for our objects…*”
>
> *“Would it need citations to cite PIDs? Hard to get people to cite proper reference already!*”
>
> *“Archives aren’t reference in publications consistently, so it seems like a challenge to link papers to catalogues.*”

It is clear that while tools to illustrate the benefits are vital, a lot of groundwork needs to be done before those benefits can be realised. The responses to that demonstrator also attested to the diversity within the sector, the issues the heritage sector has with getting users to cite their collections accurately without using PIDs, whereas this is much more mature in the earth sciences sector (Callaghan, 2012).

Part of the groundwork that’s required also appears to be cross sector support, join up or ways to integrate into cross-sector activity.

> *“exploring how to develop wider framework and integrate this internally and externally”* - A response to the question ‘Any comments on your organisation’s PID maturity?’.
>
> *“Unanimity with other cooperating institutions about identifying metadata”* - A response to the question ‘What do you or your organisation need to drive PID adoption at your organisation?’
>
> *“A place to form groups of people working in the same problem”* - A response to the question ‘What do you or your organisation need to drive PID adoption at your organisation?’

While these stop short of a call for a national policy of governance around heritage use of PIDs, the question of a national approach is one that needs to be investigated in more detail.

All webinar materials and Mentimeter results are available to download, see https://doi.org/10.23636/1174 for the launch webinar and https://doi.org/10.23636/1189 for the second webinar.

Case studies

The project has completed two case studies, one for the British Library and another for the National Gallery. These case studies had a common theme of adopting a middleware solution that connects the organisation’s core collection or library management systems with a tool to

generate identifiers. These middleware-generated PIDs are then incorporated back into the core systems and used as an identifier across institutional infrastructure.

Many collection and library management systems do not natively support PIDs and even fewer generate them. Yet these systems are core strategic infrastructure which are a major undertaking to replace. Therefore by integrating a separate lightweight system, it enabled a faster uptake of PIDs at less overhead. In the British Library's case the lightweight system's simplicity has been a virtue meaning it has required little maintenance since its first implementation in 2012. It is possible that a recommendation including some middleware type solution, may meet several other organisation's needs, and this will be explored in the second phase of the project.

The National Gallery's case study illustrated how persistent identifiers can address internal use cases of locating items as much as external use cases. That implementation will also increase efficiency by providing a mechanism that external services such as Art UK and Google Arts and Culture could in future use to automate update of harvested information.

Literature Review

With globally unique, resolvable, and persistent identifiers having been available for citation of research outputs for 20 years (Lammey, 2014; Pentz, 2001), there is a wealth of research in support of the benefits of PIDs. These benefits range from:

- Creation of trusted citations and research reproducibility (PIDs are a requirement for the FAIR principles that ensure research is Findable, Accessible, Interoperable and Reusable (Stodden & Miguez, 2014; Wilkinson et al., 2016)
- The permanence of links to information sources (Kunze, 2003)
- Increased ability and efficiency to generate usage metrics (Moss & Lyle, 2018; National Information Standards Organization, 2016)
- Improved information exchange (Meadows et al., 2019; Meadows & Haak, 2018) and management (Wang et al., 2017).

It is clear that the focus on the use of PIDs, and so the available literature, has very much centred on the sciences and social sciences. There are fewer examples that specifically deal with the use of PIDs within cultural and heritage organisations, and while the focus or relative importance of the benefits might differ, the broad scope of benefits is common across disciplines.

The Linked Heritage[8] project articulated specific requirements for PIDs but also for cultural heritage organisations in using PIDs (McKenna & Fokke, 2013). Their requirements and considerations are themselves aggregated from the literature available. Similar to our interim findings, there is an emphasis in their recommendations on setting up robust institutional policy around PIDs, and in articulating the benefits of PIDs in a form that can be used to make strong internal business cases for implementation. Their benefits are outlined as being direct benefits (such as more efficient information retrieval and management) and indirect (for example,

[8] https://www.linkedheritage.eu/index.php?en/138/about

improved ability to push information out beyond institutional boundaries, and easier management of intellectual property).

The benefit of PIDs for linking across collections is inferred by the recommendation made by the European Commission that parties involved in the Europeana project[9] ensure:

> *"the use of common digitisation standards defined by Europeana in collaboration with the cultural institutions in order to achieve interoperability of the digitised material at European level,* ***as well as the systematic use of permanent identifiers****;"* (The European Commission, 2011)

Further demonstrations of how linking content with PIDs helps to understand that content, its origins and its further uses are also available. The FREYA project[10] has built tools that enable the visualisation of connected PIDs in 'PID Graphs' (Fenner, 2020)[11], as well implementations of connected PIDs for improved discovery (Dohna, Tina et al., 2019; Lavasa et al., 2019). Consistent with comments in project webinars, the power of these tools lies in as many organisations having PIDs that can connect their content as possible.

Generally, identifiers are seen as a necessity for trustworthy data aggregation and exchange (Wickett et al., 2014). Indeed, identifiers are increasingly just another one of the elements of metadata that should be created for digital objects (including digital representations of physical objects or their digital metadata files), for example at digitisation (Blagoderov et al., 2012; Fabian & Schreiber, 2014).

The long term utility of PIDs supports preservation, which is a key benefit for cultural heritage organisations. Digital Preservation Europe (DPE) promotes cooperation beyond national boundaries and works to raise the profile of digital preservation. DPE provides several briefing papers on persistent identifiers that cover aspects such as identifier interoperability (Paskin, 2008) and identifiers for audiovisual materials (Wallaszkovits & Liebl, 2009). The most useful work for heritage identifiers is their paper on Persistent Identifiers for Cultural Heritage (Bellini et al., 2008). Further work from the APARSEN project also highlighted the benefits of PIDs for Linked Open Data and preservation of digital heritage materials (Solodovnik & Budroni, 2015).

Examples focussed on the benefits of PID use for heritage organisations are out there, but as they are less ubiquitous, and so it's understandable why responses to the survey and webinars call for better demonstrations of these benefits. This is somewhat a vicious circle, since demonstrable benefits need some implementations to go ahead. We will actively build these existing pieces of work into our supporting information for the sector.

There have been a few recent publications including Koster's article which provides a survey of various PID systems and the requirements for PIDs for cultural heritage objects (Koster, 2020). There are several blog posts also describing specific implementations of persistent identifiers, e.g. relating to Project Omega at The National Archives (Project Omega, 2020). Earth and Natural Sciences can be seen to have some similar requirements and barriers to

[9] https://www.europeana.eu/en
[10] https://www.project-freya.eu/en
[11] e.g. https://commons.datacite.org/

the cultural heritage sector, especially relating to the scale of the collections held and there is some literature relating to Natural Science Identifiers and International Standard Geo Numbers (Lehnert et al., 2019).[12]

From the literature it became apparent that there have been efforts to integrate with PIDs with some commercial collection management systems in the past, as mentioned in the webinar findings. For example, the Dutch Digital Heritage Network conducted a project which offered support to implement PIDs within heritage organisations and several vendors developed integrations for various organisations in the Netherlands (van Veenendaal et al., 2017). For example:

- Picturae have made Handles available in the Memorix software used by Noord-Hollands Archief, Archief Eemland amongst others
- DeventIT have implemented Handles in Atlantis used by Centraal Bureau voor Statistick amongst others
- Cit, distributor of The Museum System, have implemented handles with Wereldculturen group, which includes institutions such as the Tropenmuseum and the Afrika Museum as users.

An updated list from December 2019 is available through the PID Wijzer website and indicates 19 institutions in the Netherlands are utilising Handles for collection items via SURFSara.[13]

The project will continue to add literature to the HeritagePIDs Zotero group.[14]

Conclusions

Our initial work within the first eight months of the Persistent Identifiers as IRO Infrastructure project indicates that there is an awareness of identifiers within some cultural heritage organisations, particularly independent research organisations. This is a positive finding as awareness of identifiers in itself is not the barrier to adoption we thought it may be. The biggest barrier to PID adoption appears to be a lack of ability to articulate the full value proposition of globally resolvable persistent identifiers for collection items. Without this value being clear, organisations will struggle to make the business case for implementation.

Where there is uptake of identifiers, this is not necessarily of identifiers that have third party governance or guarantees of persistence. It became clear from both of the case studies that the drivers for implementing PIDs were to address internal management of content as much as external use. For example, the British Library implemented ARKs to assist in managing the increase in digital and digitised collection items as a result of legislation requiring deposit of non-print material at the Library (The Legal Deposit Libraries (Non-Print Works) Regulations, 2013). It may be that identifiers that are not globally unique and resolvable still support organisational requirements, and that may be enough for the meantime. But a full exploration

[12] https://dissco.tech/2020/05/28/natural-science-identifiers-cetaf-stable-identifiers/
[13] https://www.pidwijzer.nl/en/persistent-identifiers-in-cultural-heritage-and-research-institutes-in-the-netherlands
[14] https://www.zotero.org/groups/2470432/heritagepids/

of the additional benefits that externally governed PIDs would provide is needed to support those business cases.

Recommendation 1: The value proposition of PIDs needs to be fully articulated, and addressed directly to decision makers within cultural heritage institutions.

The project will continue to build on its findings to this end and generate materials specifically for decision makers. Our work so far has provided some clarity on how to articulate PID benefits specific to the heritage sector. Those benefits can be grouped around the themes of efficiency, trust and interoperability. Efficiency relates to the time saved by the improved management information and reporting capability that comes from having PIDs assigned to collection items. Interoperability is linked with efficiency in that it makes manual tasks easier or automates them. For example, the National Gallery's PID implementation will pave the way for information with aggregation services such as Art UK and Google Arts and Culture to be updated automatically.

In support of interoperability, the demonstrator tool developed by the project, Annotate It, has shown the benefits which could be achieved through the use of persistent identifiers to track citation and mention of collection items in any online publications. This speaks to the core concept of our project for Towards a National Collection: That the interoperability benefits of PIDs are the key factor in their use for dissolving barriers between heritage collections. They can link concepts (i.e. subjects, people and places within the metadata and authority files) common to collection items across institutional boundaries, and facilitate access - either by humans or machines - to content and metadata.

Recommendation 2a: In support of choosing appropriate identifiers, the project will develop a description of broad institutional requirements, defined to a set of 4-5 levels of complexity and matched up to the features of various identifiers.

This idea is similar to the 'Five stars of Open Data'[15] approach, and inspired by the PID wijzer tool, this would aim to explain how each additional requirement or level of requirements placed on a PID comes with benefits and costs to the organisation. It would then start to point users in the direction of the PID technologies that are available to them that match requirements at each level. This advice will build on the very early definitions that we have available in this report's List of identifiers, but the full exploration and set of definitions will be delivered in 2021.

Recommendation 2b: Subsequently, the project will create guidance which will outline practical steps to help organisations move between these levels and work with PIDs that support more complex use cases.

The findings of the survey and webinars indicated a lack of decision making capacity within organisations and this will help users identify and address their own organisation's needs.

In smaller institutions a particular barrier is the ability of commercially procured software to assign or manage PIDs. Until solutions are readily available, it will be challenging for many organisations to implement any type of PID solution. The example of the Dutch Heritage

[15] https://5stardata.info/en/

Network's project indicates that overcoming this can help drive adoption across the sector. Vendors will add support for PIDs into their systems at their own rate, but we anticipate that our future projects outputs (for example further case studies and elaboration of PID requirements) will continue to feed into that process.

> *"Produce recommendations for suppliers and providers of library systems etc too"* - A response to a request for suggestions of project outputs from the launch webinar.

Recommendation 3: We strongly recommend that heritage organisations and IROs start to work with their system suppliers to ensure systems meet their PID-based requirements (as defined in Recommendation 2A).

As well as a lack of middleware, almost no collections have an online presence for the entirety of the collection. This is vital for PIDs that can link users to actual content. This barrier also appeared to be linked to an issue of online collections being viewed as of lesser value than physical items or that research is not a critical aspect of an organisation's mission. It is unclear whether the COVID-19 situation throughout 2020 will have changed organisational attitudes around making collections available online, but it would be possible to associate creation of PIDs as part of the digitisation and publishing workflow (Fabian & Schreiber, 2014), but there would be costs associated. Certainly these would be less than assigning the identifiers retrospectively, and a clear articulation of the benefits to enabling long-term collection management, citation and metrics could support the business case for that additional cost (see Recommendation 1).

The lack of community consensus appears to have a paralysing effect on some organisations in terms of implementing PIDs, due to lack of clarity on what solutions meet their and their community's needs. As mentioned above, beyond awareness there is a need for agreement and education about PIDs to aid decision making within organisations. Beyond the lifetime of our project, heritage organisations will need to make use of the materials we will produce to generate their own policies on persistent identifiers, and actually move on to implementation. That effort will also support work that needs to be done on understanding the costs of implementation. While there are low cost options available, staff resources are always required. We are still to gather further evidence on the costs of implementation within the heritage sector, and what costs we can collate in the next 18 months will be based on a very small sample size. For the sector to understand the full costs - and benefits - of PID implementation, we will need a larger and more diverse sample of organisations to begin the journey, and this is a clear question for later phases of the Towards a National Collection Programme.

Recommendation 4: The project and IROs to continue to gather cost information on PID implementation, in particular across a more diverse sample of organisations.

Developing further cost models will allow the sector to understand whether these activities are scalable and sustainable. At this point in the project, it is not possible to answer whether these implementations are sustainable at a national level and whether costs can be borne by existing funding or whether further infrastructural support will be required. In the case studies so far, we found that ongoing costs are dependent on the solution in place, but are minimal compared

to set up. Large-scale adoption of PIDs has benefits for research into the building of collections as well as their content. If as a sector we want large-scale adoption, then this is infrastructure that may require founding investment.

For scalability, there are a few observations which provide a starting point for investigation. The implementations that exist or are under way have an awareness of interoperability in their design and seem to be informed by linked open data approaches particularly. The fact that interoperability has been designed into these solutions is promising for their scalability, however this has not been investigated in detail at this point in the project. Another aspect to address this question, which will be investigated in the remainder of the project, is who can benefit from this interoperability that is being created and how sustainable those aggregators and indexers are. There needs to be a demonstrable benefit for these implementations to make them scale.

The Annotate It demonstrator tool is manually populated at this point, and further investigation would be required to see how scalable it could be with large amounts of data. In addition, this demonstration received feedback that it was already difficult to track citation of materials held in archives and museum objects.

Recommendation 5: This project to offer some additional guidance to staff working with collections on how citation practices for heritage artefacts could be enhanced with the use of identifiers.

Our findings indicate that while demonstrators illustrate the value of persistent identifiers, many of those sorts of tools can feel very far removed from the day to day work in heritage organisations. In addition, the diversity of practice across the sector may indicate that adoption by end users e.g. of specimen collections vs archives, may progress at different rates.

Recommendation 6: Sector-wide governance and policies for PIDs should be investigated as an option to encourage uptake and to have a coherent approach to implementations and use of PIDs. The PIDs as IRO Infrastructure project will begin this work and make further recommendations on a sector-wide approach, but these will need to be tested and refined by Towards a National Collection's future Discovery Projects.

Webinar responses indicated an appetite for strategic join up across the sector. Strategic guidance to help decision makers within organisations would go some way to facilitating implementations, but some of the wider benefits of PIDs will be enhanced with a coherent approach. A national-level approach could also address the issues around locally resolvable PID types and their potential lack of long-term persistence. There also needs to be a tangible service which organisations can include their collections within in order to drive adoption. It's unclear whether a very narrow approach to a national strategy (for instance selecting a single type of identifier that all UK institutions should use for their content) would be successful. Attempts have been made, for example with the PersID project in 2011, but a statement of common principles such as Den Haag Manifesto (Treolar, 2011) may be a more appropriate starting point.

The additional case studies that are planned for the remainder of the project will drill deeper into requirements from some organisations who do not yet have PIDs to identify a broader range of use cases. From the case studies it is evident that light weight add-ons which integrate with existing systems work and can be in place within relatively short timeframes.

Our findings provide some clear direction for Towards a National Collection's future Discovery Projects, specifically:

- Recommendation 7: More IROs, HEIs and heritage organisations should implement policies on use of PIDs to support linking of items and their metadata across institutional boundaries, and identify a minimum technical passive provision for PIDs that future-proofs new tools and systems for their use.
- Recommendation 8: Where key strategic systems cannot be easily reworked for PID use:
    - A: the recommendation is for lightweight add on software which can be integrated alongside existing systems.
    - This may still be beyond the reach of smaller organisations with little or no technical capacity, and so shared infrastructure approaches in support of such organisations should be explored.
- The Towards a National Collection programme, in collaboration with IROs, heritage organisations, higher education institutions and future projects, should explore a sector-wide approach to identifiers that enables the benefits of wider take up, while respecting the diverse internal management needs and processes of organisations.

List of identifiers

This list of identifiers isn't exhaustive, and only covers those mentioned within the work so far. The aim is to define them for colleagues who are new to persistent identifiers, and give a very high-level indication of functionality as relates to our definition of 'persistent identifiers' as those that are:

- Globally unique: This means that there are no clashes with the identifier anywhere on the web. This also means the identifier can only be used for the same logical content. If the same identifier were used for two different items, that would effectively be a clash.
- Actionable: This means that it can be formatted as a link that users can click to take them to a resource or information about a resource. This is achieved with a resolution URL, e.g. 'https://purl.org/' can be added to 'dc/elements/1.1/' to make the actionable identifier https://purl.org/dc/elements/1.1.
- Persistent: The identifier is managed to remain unique and actionable for the long term. Persistence may be guaranteed by the policies and service agreements of the organisation creating the PIDs, or may be underwritten by third party agreements and governance.

There is room for nuance within the definition of each of these for each of the elements above, particularly around persistence, which is why this is not a simple 'TRUE/FALSE' assignment in the table. But also it should be remembered that there is room for human error - for example a well managed rule within a PID system for uniqueness may accidentally be broken by a user.

Identifier Name	**Description**	**Is globally unique?**	**Is actionable?**	**How is persistence guaranteed?**	**Links**
ARKs	Archival Resource Key, a PID infrastructure for digital objects, that can be locally installed	TRUE	TRUE	Assigning organisation would need to set up management for persistence	https://n2t.net/e/ark_ids.html
CETAF Stable Identifiers	Consortium of European Taxonomic Facilities scheme of identifiers for specimens	TRUE	TRUE	UNKNOWN	https://cetaf.org/cetaf-stable-identifiers
Dewey	Dewey Decimal Classification is a proprietary library classification	TRUE	FALSE	Managed by OCLC	http://www.gutenberg.org/files/12513/12513-h/12513-h.htm

Identifier Name	Description	Is globally unique?	Is actionable?	How is persistence guaranteed?	Links
DOI	Digital Object Identifier, a globally governed PID infrastructure. DOIs are particularly used for research-related objects	TRUE	TRUE	Through DOI governance	https://www.doi.org/factsheets/DOIKeyFacts.html
EIDR	Entertainment Identifier Registry Association, provides universal identifiers for movie and television assets, based on DOI infrastructure	TRUE	TRUE	Through DOI governance	https://www.eidr.org/about-us/
Geonames	A database of geographical names combining data from multiple sources	TRUE	TRUE	UNKNOWN	https://www.geonames.org/about.html
Handle	A distributed PID infrastructure for digital objects, that can be locally installed	TRUE	TRUE	Assigning organisation would need to set up management for persistence	http://handle.net/index.html
ISBN	International Standard Book Number, a product identifier for text based monographic publications	TRUE	FALSE	Managed by The International ISBN Agency	https://www.isbn-international.org/content/what-isbn
ISNI	International Standard Name Identifiers, for disambiguation of authors, creators and copyright holders (as individuals and as organisations)	TRUE	TRUE	Managed by OCLC	https://isni.org/page/what-is-isni/
LSID	Life Science Identifiers, for data resources	TRUE	TRUE	UNKNOWN	http://www.lsid.info/
IGSN	International Geo Sample Number, identifiers for physical samples used in research	TRUE	TRUE	Governance of IGSN	https://www.igsn.org/about/
MDA Codes	Managed by the Collections Trust, MDA Codes identify UK collection holding organisations and their collections	TRUE	FALSE	Managed by Collections Trust	https://collectionstrust.org.uk/resource/mda-codes

Identifier Name	Description	Is globally unique?	Is actionable?	How is persistence guaranteed?	Links
ORCID	Open Researcher and Contributor ID, for disambiguating research authors and contributors	TRUE	TRUE	Managed by ORCID	https://support.orcid.org/hc/en-us/articles/360006971053-Your-ORCID-iD-your-digital-name-identifier
ROR	Research Organization Registry, identifiers for research organisations	TRUE	TRUE	Managed by ROR	https://ror.org/about/
PURL	Persistent Uniform Resource Locator, a dereferenced and persistent URL for digital objects, managed by the Internet Archive	TRUE	TRUE	Assigning organisation would need to set up management for persistence	https://archive.org/services/purl/help
UMID	Unique Material Identifier, provides a method of identification for instances of audiovisual material and thus enables the material to be linked with its associated metadata	TRUE	UNKNOWN	UNKNOWN	https://doi.org/10.5594/SMPTE.ST330.2011
URN/URI	Uniform Resource Name / Uniform Resource Identifier, identifying objects within a given domain or namespace	TRUE	TRUE	Assigning organisation would need to set up management for persistence	https://www.w3.org/TR/uri-clarification/
UUID	Universally Unique Identifier, used to identify information in computer systems	TRUE	FALSE	Assigning organisation would need to set up management for persistence	https://doi.org/10.17487/RFC4122
VIAF	Virtual International Authority File, combines multiple authority files from libraries around the world into one service	TRUE	TRUE	Managed by OCLC	https://www.oclc.org/en/viaf.html
Wikidata	Provides human and machine readable structured information. The central storage of structured data of the Wikimedia projects	TRUE	TRUE	Managed by Wikidata	https://www.wikidata.org/wiki/Wikidata:Main_Page

References

Bellini, E., Cirinna, C., & Lunghi, M. (2008). *Briefing Paper: Persistent Identifiers for Cultural Heritage*. Digital Preservation Europe. https://www.digitalpreservationeurope.info/publications/briefs/persistent_identifiers.pdf

Blagoderov, V., Kitching, I., Livermore, L., Simonsen, T., & Smith, V. (2012). No specimen left behind: Industrial scale digitization of natural history collections. *ZooKeys*, *209*, 133–146. https://doi.org/10.3897/zookeys.209.3178

Callaghan, S. (2012). Data Citation in the Earth and Physical Sciences. In *For Attribution—Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop* (p. 13564). National Academies Press. https://doi.org/10.17226/13564

Concordat Working Group, & others. (2016). *Concordat on Open Research Data*. https://www.ukri.org/funding/information-for-award-holders/data-policy/

Dohna, Tina, Koop-Jakobsen, Ketil, Schindler, Uwe, Lavasa, Artemis, Fokianos, Pamfilos, Dallmeier-Tiessen, Sünje, Bunakov, Vasily, Madden, Frances, & Ferguson, Christine. (2019). *Deliverable 4.3 Using Advanced PIDGraph Functionality inPilot Applications*. https://doi.org/10.5281/ZENODO.4066842

Fabian, C., & Schreiber, C. (2014). Piloting a National Programme for the Digitization of Medieval Manuscripts in Germany. *LIBER Quarterly*, *24*(1), 2. https://doi.org/10.18352/lq.9265

Fenner, M. (2020). *The PID Graph in FREYA (additional project report)*. https://doi.org/10.5281/ZENODO.4028383

Groom, Q., Hyam, R., & Güntsch, A. (2017). Stable identifiers for collection specimens. *Nature*, *546*(7656), 33–33. https://doi.org/10.1038/546033d

Koster, L. (2020). Persistent identifiers for heritage objects. *Code4lib*, *47*. https://journal.code4lib.org/articles/14978

Kroon, R. (2014). Bringing Order to Digital Identifiers. *M&E Journal*, *2014–2015*(Q4), 148–150.

Kunze, J. (2003). *Towards electronic persistence using ARK identifiers*. https://escholarship.org/uc/item/3bg2w3vs

Lammey, R. (2014). CrossRef developments and initiatives: An update on services for the scholarly publishing community from CrossRef. *Science Editing*, *1*(1), 13–18. https://doi.org/10.6087/kcse.2014.1.13

Lavasa, A., Dallmeier-Tiessen, S., Sandt, S. V. D., Dohna, T., Koop-Jakobsen, K., Schindler, U., Ferguson, C., McEntyre, J., Madden, F., Lambert, S., Bunakov, V., & Baars, C. (2019). *D4.2 Using the PID Graph: Provenance in Disciplinary Systems*. https://doi.org/10.5281/ZENODO.3249832

Lehnert, K., Klump, J., Wyborn, L., & Ramdeen, S. (2019). Persistent, Global, Unique: The three key requirements for a trusted identifier system for physical samples. *Biodiversity Information Science and Standards*, *3*, e37334. https://doi.org/10.3897/biss.3.37334

McKenna, G., & Fokke, C. (2013). *D2.2 State of the art report on persistent identifier standards and management tools*. https://www.linkedheritage.eu/index.php?en/215/persistent-identifiers

Meadows, A., & Haak, L. (2018). How persistent identifiers can save scientists time. *FEMS Microbiology Letters*, *365*(15). https://doi.org/10.1093/femsle/fny143

Meadows, A., Haak, L. L., & Brown, J. (2019). Persistent identifiers: The building blocks of the research information infrastructure. *Insights the UKSG Journal*, *32*, 9. https://doi.org/10.1629/uksg.457

Moss, E., & Lyle, J. (2018). *Opaque data citation: Actual citation practice and its implication for tracking data use*. http://hdl.handle.net/2027.42/142393

National Information Standards Organization. (2016). *Outputs of the NISO alternative assessment metrics project*. https://www.niso.org/publications/rp-25-2016-altmetrics

Paskin, N. (2008). *Briefing Paper: Identifier Interoperability*. Digital Preservation Europe. https://www.digitalpreservationeurope.info/publications/briefs/identifier-interoperability.pdf

Pentz, E. (2001). CrossRef: A collaborative linking network. *Issues in Science and Technology Librarianship*, *10*, F4CR5RBK. https://doi.org/10.5062/F4CR5RBK

Project Omega, T. N. A. (2020, June 3). Archival Catalogue Record Identifiers. *The National Archives Digital*. https://medium.com/the-national-archives-digital/archival-catalogue-record-identifiers-29b0a1fac9ba

Research Councils UK. (2011). *RCUK common principles on data policy*. https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/

Solodovnik, I., & Budroni, P. (2015). Preserving digital heritage: At the crossroads of Trust and Linked Open Data. *IFLA Journal*, *41*(3), 251–264. https://doi.org/10.1177/0340035215600453

Stodden, V., & Miguez, S. (2014). Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. *Journal of Open Research Software*, *2*(1), e21. https://doi.org/10.5334/jors.ay

The European Commission. (2011). Commission Recommendation of 27 October 2011 on the digitisation and online accessibility of cultural material and digital preservation (2011/711/EU). *Official Journal of the European Union*, 39–45.

The Legal Deposit Libraries (Non-Print Works) Regulations, Pub. L. No. 777 (2013). https://www.legislation.gov.uk/uksi/2013/777/contents/made

Treolar, A. (2011). *Den Haag Persistent Object Identifier – Linked Open Data Manifesto*. http://repository.jisc.ac.uk/6438/2/Den_Haag_Persistent_Object_Identifier_%E2%80%93_Linked_Open_Data.pdf

van Veenendaal, R., Ras, M., & Dangerfield, M. C. (2017, September 25). *Getting Persistent Identifiers Implemented By 'Cutting In The Middle-Man'*. iPres 2017, Kyoto, Japan. https://doi.org/10.5281/zenodo.3556530

Wallaszkovits, N., & Liebl, C. (2009). *Briefing Paper: UMID - Unique Material Identifier*. Digital Preservation Europe. https://www.digitalpreservationeurope.info/publications/briefs/UMID_Unique%20Material%20Identifier.pdf

Wang, J., Car, N., Evans, B., Gohar, K., Trenham, C., & Wyborn, L. (2017). Persistent Identifier Practice for Big Data Management at NCI. *Data Science Journal*, *16*, 20. https://doi.org/10.5334/dsj-2017-020

Wickett, K. M., Isaac, A., Doerr, M., Fenlon, K., Meghini, C., & Palmer, C. (2014). Representing Cultural Collections in Digital Aggregation and Exchange Environments. *D-Lib Magazine*, *20*(5/6). https://doi.org/10.1045/may2014-wickett

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Appendix 1 - About Persistent Identifiers as IRO Infrastructure

Museums, heritage collections and sites in the UK house at least 200 million physical and digital objects. Being able to identify these objects supports their discovery, use and curation – you cannot provide persistent or even consistent access to an item if you don't know what it is. Accession numbers are a key component in all collection and library management systems but these only cover selected objects within an individual collection. To fully realise the potential of our national collections, we need identifiers that can bring together collections across institutional boundaries.

Persistent Identifiers (PIDs) provide a long-lasting click-able link to a digital object. They are recognised by UKRI as a tool for enabling data discovery, access and citation. Supporting wider use of PIDs for collection objects, environments, specimens and related items will allow long-term, unambiguous linking of collections that will create a digital National Collection.

However, the challenges, utility and wider benefits of PIDs are not as well understood across the heritage sector as they could be.

This project will bring together best practices in the use of PIDs, building on existing work and projects. We will share expertise and provide recommendations on the approach to PIDs for colleagues in institutions across the UK heritage sector. Through a mixture of workshops, desk research and case studies, the project will answer questions such as 'What are the gaps in the existing PID landscape for heritage collections, buildings and environments?' and 'What should a PID infrastructure, strategy and governance framework look like for a unified national collection?'.

This project is a Foundation project within the AHRC funded Towards a National Collection Programme[16] project reference AH/T011092/1.

Appendix 2 - About Towards a National Collection

Towards a National Collection is a major five-year £18.9 million investment in the UK's world-renowned museums, archives, libraries and galleries. Funding is provided through UK Research and Innovation's Strategic Priorities Fund and delivered by the Arts and Humanities Research Council (AHRC). The programme will take the first steps towards creating a unified virtual 'national collection' by dissolving barriers between different collections – opening UK heritage to the world. By seizing the opportunity presented by new digital technology, it will allow researchers to formulate radically new research questions, increase visitor numbers, dramatically expand and diversify virtual access to our heritage, and bring clear economic, social and health benefits to communities across the UK. The innovation driven by the programme will maintain the UK's world leadership in digital humanities and set global standards in the field.

The Programme's main objectives are:

- to begin to dissolve barriers between different collections
- to open up collections to new cross-disciplinary and cross-collection lines of research
- to extend researcher and public access beyond the physical boundaries of their location
- to benefit a diverse range of audiences
- to be active and of benefit across the UK
- to provide clear evidence and exemplars that support enhanced funding going forward.

Aims of the Programme

The aim of the Programme is to begin to dissolve barriers between different collections, opening them up to new cross-disciplinary and cross-collection lines of research, and to extend researcher and public access beyond the physical boundaries of their location, thus directly addressing the issues related to accessibility beyond current metropolitan centres. The programme will extend across the UK including all the devolved nations, and will potentially have a global reach in terms of setting a standard for other countries building their

[16] https://tanc-ahrc.github.io/HeritagePIDs/

own collections (with the long-term potential for inter-connection between the national collections).

This Programme will have a transformative impact on:

- Digital search and cataloguing tools, technologies and methodologies, and associated issues
- Research capability, by enabling search across collections to address cross-cutting research questions which will allow UK to maintain UK leadership in cross-disciplinary research
- The heritage sector as a whole, in terms of enhancing access for researchers, and for facilitating wider and better-informed public engagement.

There are two rounds of funding calls – the Foundation Projects and the Discovery Projects[17]

Image Credits from Figure 1

Awareness Icon, Ivan Colic from Noun Project
Barrier Icon, Vectors Point from Noun Project
Document Icon, Maxim Kulikov from Noun Project

Organisation Icon, Mike Wirth from Noun Project
Painting Icon, Kiran Shastry from Noun Project
Person Icon, Made from Noun Project
Sample Icon, Becris from Noun Project
Service Icon, Oksana Latysheva from Noun Project

[17] https://www.nationalcollection.org.uk/about