



TOWARDS
A NATIONAL
COLLECTION



Arts and
Humanities
Research Council

FINAL REPORT

FOUNDATION PROJECTS

Persistent Identifiers as IRO Infrastructure

PI: Rachael Kotarski, British Library

The British Library | Royal Botanic Garden Edinburgh
The National Gallery | University of Glasgow
Victoria and Albert Museum | Science Museum Group
Natural History Museum

March 2022

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
ABSTRACT	3
AIMS AND OBJECTIVES	3
PARTNERSHIP STRUCTURE	5
STAFFING STRUCTURE	5
REVISED OVERALL PROGRAMME	6
EVENTS AND CONSULTATIONS	7
RESEARCH APPROACH	8
RESEARCH RESULTS	9
PROJECTS OUTPUTS	14
RECOMMENDATIONS FOR THE PROGRAMME.....	16
CONTACTS	19
ANNEX	20

AUTHORS

*[Jack Kirby](#) (Science Museum Group), [Rachael Kotarski](#), (British Library)
[Frances Madden](#) (British Library), [Lorna Mitchell](#) (Royal Botanic Garden Edinburgh),
[Joseph Padfield](#) (National Gallery), [Roderic Page](#) (University of Glasgow),
[Richard Palmer](#) (Victoria and Albert Museum) &
[Matt Woodburn](#) (Natural History Museum, London)*

Executive Summary

Heritage organisations in the UK house at least 200 million physical and digital objects¹. Being able to uniquely 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. Persistent Identifiers (PIDs) provide a long-lasting click-able link to a digital object, making content Findable, Accessible, Interoperable and Reusable (FAIR)² and enabling citation and metrics. Supporting wider use of PIDs in collection organisations will allow long-term, unambiguous linking that will help create a digital national collection. *Persistent Identifiers as IRO Infrastructure* brought together best practices, gathering evidence to develop an effective toolkit for the sector to make wider use of PIDs.

Through two surveys, four long-form case studies and a number of virtual workshops and events, we heard of barriers to PID use that covered awareness and buy-in from decision-makers. Guidance produced by the project has already begun to reduce these barriers. However, wider access to technical infrastructure still presents a significant barrier as discussed in the closing part of the project, with the lack of infrastructure and resource (both technical and staff-wise) for smaller organisations being the biggest issue for their participation in connected digital collections. Persistent identifiers can easily be integrated into new tools for digital management of collections, considerably lowering the costs and expertise required for PID adoption compared to adoption as a discrete and separate activity. While bought-in vendor solutions may provide PIDs, staff still need appropriate training and guidance to ensure PIDs are used to their full benefit. This shifts the costs from implementation technologies and activities, to professional development. As a result, we recommend a continuation of cost analysis on PID implementation from organisations just starting use of PIDs, in particular across a more diverse sample of organisations. Costs should consider professional development in support of better use of integrated PIDs as well as stand-alone PID implementation costs, and continued discussion with vendors will help define and start to reduce these issues. Through these discussions, the sector can start to ensure that new strategic infrastructure supports creation and use of PIDs for content, and use of PIDs in metadata. In the meantime, a light-weight middleware approach will support initial PID implementations for larger organisations.

We have seen through the existing diversity of the approaches to persistent identifiers that a sector-wide approach cannot and should not be overly prescriptive in the types of persistent identifiers that should be used. Individual organisational needs (e.g. linked data metadata approaches vs. collection identification vs. machine readability) and capacity vary considerably and determine the most appropriate identifier tool(s). There are clear sector-wide benefits to a networked approach, in terms of making connections between collections, making collections FAIR, enabling metrics and reducing management burden. But a networked approach can still be built on common principles and functionality, such as those outlined within our '*Developing Identifiers for Heritage Collections*' guidance³, without necessarily requiring all organisations to use the exact same identifier tools. Community principles can also act as a point of discussion with system

¹ Keene, S; Stevenson, A; Monti, F; (2008) Collections for people: museums' stored collections as a public resource. UCL Institute of Archaeology: London <https://discovery.ucl.ac.uk/id/eprint/13886>

² Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

³ <https://tanc-ahrc.github.io/PIDResources/>

vendors, and speak to the varied needs and use cases of the sector, to avoid approaches that implement 'PIDs for PIDs sake' and that deliver measurable benefits.

We do recommend that the framework of requirements produced by this project for the '*Developing Identifiers*' resource should be used as a starting point for the sector, as it allows a common understanding and articulation of those requirements. This leads us to recommend that a path to community adoption of the resource should be found so that it can be maintained, updated, and adapted, as use of PIDs in the community widens.

We welcome continued contributions to the resource as a community effort, and call on colleagues to continue to feed into development of the resource either: as a user, by feeding in new requirements and questions; as a contributor, by issuing pull requests to make improvements and additions to the toolkit; or as an adapter, who can fork and tailor the resource for their specific community.

Final recommendations

- The TaNC programme, in collaboration with IROs, heritage organisations, higher education institutions and future projects, should define sector-wide principles for an approach to identifiers built on common principles, functionality, and use cases such as those outlined within '*Developing Identifiers*'. This approach does not necessarily require all organisations to use the exact same identifier, and so enables the benefits of wider take up, while respecting the diverse management needs, processes, and resource constraints of organisations, and avoiding 'PIDs for PIDs sake'
- We strongly recommend that heritage organisations start to work with their system suppliers to ensure systems meet their PID-based requirements and community principles
- The TaNC programme should look at opportunities to continue to gather cost information on PID implementation from organisations just starting use of PIDs, in particular across a more diverse sample of organisations. Costs should consider professional development in support of better use of integrated PIDs as well as stand-alone PID implementation
- 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
- Where key strategic systems cannot be easily reworked for PID use:
 - A: lightweight add-on software 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.
- The guidance within '*Developing Identifiers*' should be used as a common starting point for the sector, allowing a common understanding and articulation of requirements
- A path to community adoption of the resource will allow it to be maintained, updated, and adapted as use of persistent identifiers in the community evolves, and we call on the community to use, contribute to, and adapt the resource.

Abstract

Heritage organisations in the UK house at least 200 million physical and digital objects⁴. Being able to uniquely 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 to link together collections across institutional boundaries.

Persistent Identifiers (PIDs) provide a long-lasting, click-able link to a digital object, recognised by UKRI as a tool for making content Findable, Accessible, Interoperable and Reusable (FAIR)⁵ and enabling citation and metrics. Supporting wider use of PIDs for collection objects, environments, specimens and related items will allow long-term, unambiguous linking that will create a digital national collection. However, the challenges, utility and wider benefits of PIDs are not well understood across the heritage sector.

The project brought together best practices in the use of PIDs, building on existing work and projects. Through a mixture of workshops, surveys, desk research and case studies, the project gathered evidence to develop an effective toolkit for the sector to make wider use of PIDs and provided recommendations on an approach to PIDs for colleagues and institutions across UK heritage.

Aims and Objectives

The project aimed to increase the uptake and use of PIDs for heritage collections so that they can serve as a foundational infrastructure for drawing together the national collection, delivering innovation by adopting a cross-disciplinary and cross-collections approach to the use of an existing technology.

We aimed to gather evidence on the organisational and cultural barriers that impact on the adoption of PIDs in the sector, and how to overcome these to realise the additional benefits of PID use across collections, such as improved usage metrics; network analysis of collections; and improved links to enriched or additional information and new perspectives on collections.

We also aimed to produce a suite of resources including the final project report, videos and a demonstrator site that will provide ongoing support for the adoption of PIDs across the sector. These resources are available under licences that will encourage their sharing and re-use and contribute to the sustainability of the project outputs.

Questions posed by the project include:

- What are the benefits and implications of assigning and making use of globally unique and interoperable PIDs across national collections?
- What are the barriers to the wide scale adoption of PIDs within heritage institutions?

⁴ Keene, S; Stevenson, A; Monti, F; (2008) Collections for people: museums' stored collections as a public resource. UCL Institute of Archaeology: London <https://discovery.ucl.ac.uk/id/eprint/13886>

⁵ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

- What are the gaps in the existing PID landscape for cultural heritage?
- Are existing PID implementations scalable and sustainable at national level?
- What should a PID infrastructure, strategy and governance framework look like for a unified UK national collection?
- What should the strategic plan be for PIDs as part of a future UK collections research infrastructure?

By working across sectors at local, regional and national levels we aimed to identify key requirements for the use of PIDs within cultural heritage in the broadest sense, and map those requirements to the existing PID landscape to develop recommendations that can be promoted to government, funders and other sectors.

We sought to understand the barriers to applying PIDs to heritage collections and find a path to overcome them. With a coherent set of recommendations on PID implementation for heritage, we hoped to bring about a common approach that works with the diverse approaches to cataloguing found across the sector. In this way we aimed to deliver solutions that are scalable and provide sufficient benefit to users to ensure widespread adoption and long-term support.

With case studies, we intended to demonstrate the benefits of PIDs available to potential new users, to encourage and support take-up of the project recommendations. As we uncovered these benefits, we also wanted to demonstrate how PIDs can facilitate transparent and reproducible research in all domains, and the collection of metrics on collection use to evidence decision making (e.g., on investment in digitisation).

Partnership structure

The project partnership was formed of seven organisations. Six of the partners represent heritage Independent Research Organisations (IROs) and the seventh is the University of Glasgow.

The project was led by the British Library, who was responsible for the delivery of the project, in collaboration with the co-investigators and project partners. The British Library was responsible for project management and reporting; budgeting and funding distribution; and for recruitment of the Research Associate.

The co-investigators were staff from the Royal Botanic Garden Edinburgh (RBGE), National Gallery and University of Glasgow. Co-investigators were responsible for working with the Principal Investigator (PI) and Research Associate to feed in their organisational and technical experience to the development and use of PIDs in their respective collections, with a particular focus on the organisational challenges to PID implementation and adoption at scale. They also worked on the landscape analysis and technical framework of existing PID use and its future development across their sector as a whole, in line with project aims. We hoped that RBGE would be able to host at least one physical workshop for IROs and heritage sector stakeholders across Scotland before the end of the project, but COVID restrictions prevented this, and so we focussed on virtual conversations. The National Gallery liaised with the PI on a number of concurrent EU H2020 funded projects exploring the use of PIDs within the documentation and dissemination of Heritage Science research data and ensured the project outputs are of relevance and mutually supportive to EU developments. The University of Glasgow delivered the PID demonstrator and acted as a 'critical friend', contributing expertise and insight to consultations on PID implementation and adoption in a variety of contexts, particularly relating to user expectations for access, discovery and use of collections enabled through PID infrastructures. They also reviewed and contributed to draft reports and participated in workshops.

Project Partners were the Victoria and Albert Museum (V&A), Science Museum Group (SMG) and Natural History Museum (NHM). They each worked with the Research Associate to feed in their organisational experiences and requirements for PIDs; attended planned workshops; and reviewed the final recommendations prior to release. Science Museum Group were due to host a workshop in the North of England, to facilitate a geographic spread for the workshop series as a whole, but when this had to be moved to a virtual format, we held a joint event with the Towards a National Collection Heritage Connector project. The NHM was a vital link between the project and the EU-funded DiSSCo Project.

Staffing structure

Staff carrying out the responsibilities of the British Library were Rachael Kotarski (Head of Research Infrastructure Services as the Principal Investigator), along with the Research Associate, Frances Madden.

Co-investigators were Lorna Mitchell (Head of Library Services, RBGE); Joseph Padfield (Principal Scientist, National Gallery); and Rod Page (Professor of Taxonomy, University of Glasgow).

Staff from project partners responsible for their respective contributions were Richard Palmer (Senior Web Developer, V&A); Jack Kirby (Associate Director of Collections Services, SMG); and Matt Woodburn (Science Data Architect, NHM).

Revised overall programme

Summary	2020										2021										2022		
	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb
Webinar #1																							
Survey #1 Launched																							
Webinar #2 ⁶																							
Case study #1 – British Library																							
Case study #2 – National Gallery																							
Early Findings																							
Demonstrator tool: in beta																							
Project webinar #3																							
Case study #3 – NHM																							
Project webinar #4 ⁷																							
Case study #4 – RBGE																							
Video summarising best practices																							
Survey #2																							
Project webinar #5 ⁸																							
Final event																							
Final report																							

⁶ Based on interim survey results

⁷ Summary of first year of project, consultation of Developing Identifiers for Heritage Collections proposal & collections proposal

⁸ Joint with other foundation project (IIIF)

Events and consultations

Event	Date	Outputs	Attendees
Digitisation and the State-of-the-art(world)	05 March 2020	Event summary Recordings	70*
Project launch webinar (including Heritage Connector presentation)	06 April 2020	Materials and recording	123
Research Data Alliance PID Interest Group	09 April 2020	Materials and recording	30*
Project survey #1	28 May 2020	Summary data	66
DataCite Summer Client Meeting	15 July 2020		34*
Identifiers in Heritage Collections - how embedded are they?	17 July 2020	Materials and recording	93
National Gallery Scientific Consultative Group	23 November 2020		21
PIDapalooza	27 January 2021	Recording	122
Foundation project webinar: Persistent Identifiers and Locating a National Collection	19 February 2021	Recording	Unknown
Developing Identifiers for Heritage Collections	21 April 2021	Recording	98*
Developing Identifiers for Heritage Collections - consultation, testing	23 July 2021		514 users
LD4 Conference on Linked Data	21 July 2021	Poster	Unknown
Project survey #2	04 October 2021	Summary data	47
Persistent Identifiers in IIF	26 October 2021	Materials and recording	110*
Linked Pasts 7 Conference	14 December 2021	Poster	
Final event	18 January 2022	Materials and recording	47
Twitter	N/A		273 followers

* Figures are approximate

Research approach

The aims and objectives were met by bringing together the holders and curators of heritage collections (from across the UK as well as select international organisations) to uncover the barriers as well as the existing best practices in the application of PIDs to collections. An initial literature review as reported in the Early Findings⁹ showed that 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 highlight the specific benefits of PID use to the sector.

The most important new literature to emerge during the project is a significant cost-benefit analysis on PID adoption in UK Higher Education¹⁰. This work focused on the ability of PIDs to ease ‘pain points’ in information exchange workflows. Based purely on staff effort savings in re-keying information, estimated sector-wide benefits (including cost saving) of £5.67M if 85% adoption can be achieved within five years. Adoption is key, as high adoption accrues greater network benefits for data sharing and connectivity. These figures may not be directly comparable to UK heritage, as there is not a complete overlap in benefits (collection management activities for instance are not reflected). However, the focus on data exchange is key when discussing the network benefits for connected collections as part of Towards a National Collection.

Virtual events, changed from in-person workshops, were supported by the findings of this desk-based research to communicate the best practice identified in different sectors. In-depth analysis in the form of case studies of project partner organisations use and implementation of PIDs was carried out. The findings of this work were discussed and explored, to create a set of recommendations that have been embedded in the resources to support take-up of identifiers across the sector.

Additional steps in the research approach were the inclusion of a sector-wide survey at the beginning and end of the project, allowing us to perform an early identification of current status and community needs. PIDs as IRO Infrastructure ‘Survey 1’ was launched on 28 May 2020. A long response timeframe was given to accommodate returns from staff who may have been on furlough due to the COVID-19 pandemic. ‘Survey 2’ ran from 4 October 2021. The survey was designed with a series of benchmarking questions on the experience and familiarity with PIDs within the sector. These provided us with a way to understand the progress made during the short project.

We brought together best practice in the use of PIDs from a collection perspective, building on existing IRO work and expertise developed through research projects such as FREYA¹¹ and DiSSCo¹² (explored in the NHM case study). Ultimately, we provided a framework of recommendations on the approach to PIDs for colleagues across the UK heritage sector.

⁹ Kotarski, R., Kirby, J., Madden, F., Mitchell, L., Padfield, J., Page, R., Palmer, R., & Woodburn, M. (2020). PIDs as IRO Infrastructure - Early Findings. British Library <https://doi.org/10.23636/1214>

¹⁰ Brown, J., Jones, P., Meadows, A., Murphy, F., & Clayton, P. (2021). UK PID Consortium: Cost-Benefit Analysis (Version 1.0). Zenodo. <https://doi.org/10.5281/ZENODO.4772627>

¹¹ <https://www.project-freya.eu/en>

¹² <https://www.dissco.eu/>

Research results

Surveys

A more detailed discussion of the survey results is given in Annex, and anonymised data is available¹³. Survey 2 attracted slightly fewer responses, and at least 40% of the respondents had not completed Survey 1. However, the organisations represented amongst the responses indicate that the project has reached its intended audience.

Results suggest an increasing awareness of PIDs, just over 50% of Survey 2 respondents said they had learned something new about PIDs in the last 18 months. The results indicate that incremental progress is being made, albeit with a large amount of work to be done to embed PIDs fully in organisational GLAM (galleries, libraries, archives and museums) workflows. Given the small sample sizes of both surveys and the small overlap between the two, it is impossible to make concrete assertions about any long-term change in attitudes.

There is, however, a slight increase in the use of some persistent identifiers over the period of the project. 30 of the 47 respondents work for an organisation that uses persistent identifiers, implying that the survey represents an atypical sample of heritage professionals.

Barriers inevitably still remain, given the short duration of this work. Survey 1 highlighted 'awareness' as the primary barrier to wider use of PIDs, but by Survey 2, 'resources' and 'technology' are the primary barriers respondents feel need to be overcome (Figure 1 & 2).

Many free text responses in Survey 2 alluded to resource constraints, especially for smaller organisations. While the project has communicated about the need for PIDs and their utility, the sector has not made much progress in moving toward implementation. However, the percentage of respondents who said they were implementing or planning implementation had slightly increased, up from 46% to 50%.

Despite increasing awareness, the benefits of PIDs to decision makers within the sector are still unclear. This may require a different approach from the broader community awareness raising that we have been able to incite during the project. The final project videos¹⁴ were produced to support this particular effort, although we have not had time to measure their impact.

One key recommendation is on engagement with system suppliers. We engaged with some suppliers that offer support for linked open data, and we have reported on projects in the Netherlands to integrate PIDs with system suppliers. Integration in vendor systems may make longer-term uptake and wider adoption cheaper¹⁵. Where the project has demonstrated success is in the materials produced to support the community. While not all of the respondents had yet used a project output, those that had had all found them useful, indicating the project has met user needs (Figure 3).

Given the relatively short interval between the two case studies, it is perhaps not surprising that the changes seen between the results are relatively modest. In light of the COVID-19 pandemic and the often long

¹³ Kotarski, R., & Madden, F. (2021). Persistent Identifiers as IRO Infrastructure: Survey 2 Data [Data set]. British Library. <https://doi.org/10.23636/J95G-8446>

¹⁴ British Library. (2021). Persistent Identifiers in Cultural Heritage Collections. British Library. <https://doi.org/10.23636/13KA-JE14>

¹⁵ Stevenson, J (2021). Thoughts on the Heritage PIDs Project. Archives Hub Blog. <https://blog.archiveshub.jisc.ac.uk/2022/01/18/thoughts-on-the-heritage-pids-project/>

timescales in which cultural institutions work, this lack of perceived progress in actual use of PIDs and their implementations is to be expected.

Demonstrator tool

The Annotate It! tool shows how PIDs can be used to understand where collections connect to research¹⁶. The significance of this work is that neither the institutions managing collections, nor the publishers hosting research 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. When collection objects have PIDs, and those PIDs are cited, then we can demonstrate links between collection objects and their use by the academic community.

Case studies

The project has completed four case studies of PID use at project partners, covering the British Library, National Gallery, Natural History Museum and Royal Botanic Garden Edinburgh. Key findings from the first two case studies emphasise an approach to implementing PIDs using lightweight middle-ware, which allows uptake alongside critical strategic infrastructure. The latter case studies focus on natural history collections adopting 'stable' identifiers. These are managed in-house, which may have lower costs but require staff technical skills to maintain, and lack the external governance to support persistence.

'Mini case studies' were produced as short form examples of different aspects of PIDs for heritage collections including the perspective of an aggregator - the Archives Hub¹⁷. These have also allowed us to connect to the broader TaNC Foundation projects.

Developing Identifiers for Heritage Collections

This resource¹⁸ was developed in line with the findings of Survey 1 and illustrates the requirements of different PID implementations at differing levels of complexity. It includes guidance to help institutions with implementation, including how to ensure persistence, how to estimate costs, and how to encourage citation of collection items using PIDs. The draft resource was launched in April 2021. Following feedback and testing with the community, the resource was revised, reformatted and further guidance added.

"We found the project's compliance levels a useful model" - Feedback on outputs from Survey 2. 'Compliance levels' are part of the '*Developing Identifiers*' resource.

¹⁶ Documentation can be accessed via <https://pid-demonstrator.herokuapp.com/>

¹⁷ <https://archiveshub.jisc.ac.uk/>

¹⁸ <https://tanc-ahrc.github.io/PIDResources/> (<https://doi.org/10.5281/zenodo.5205757>)

Figure 1: Responses to Survey 1 on barriers to PID adoption seen within respondents' organisation or community

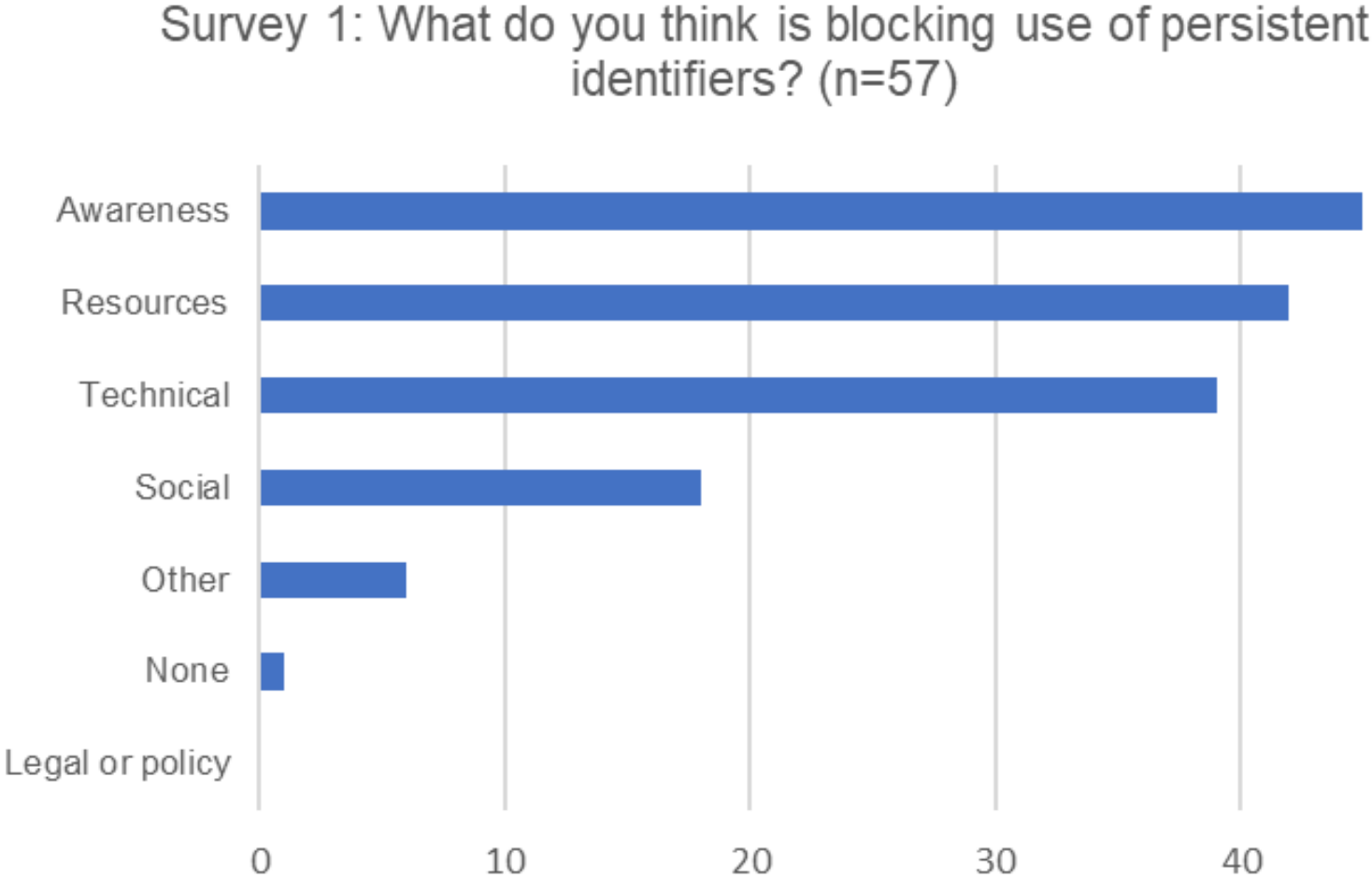


Figure 2: Responses to the same questions from Survey 2

Survey 2: What do you think is blocking use of persistent identifiers? (n=35)

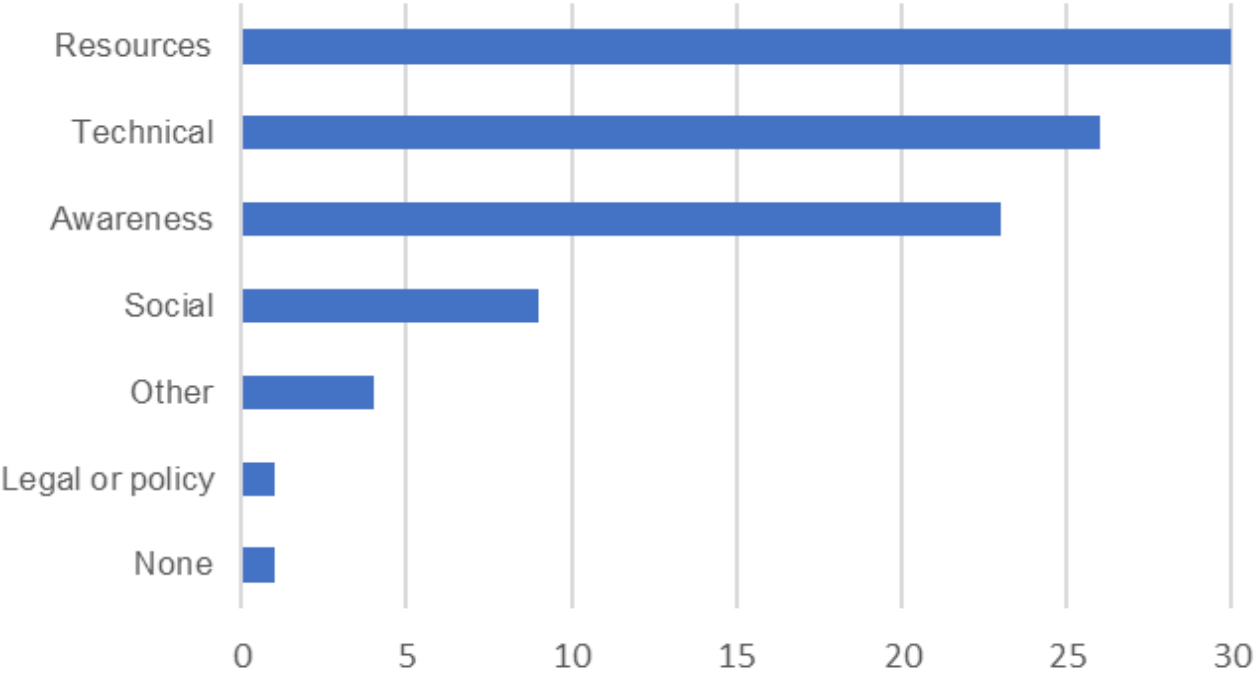
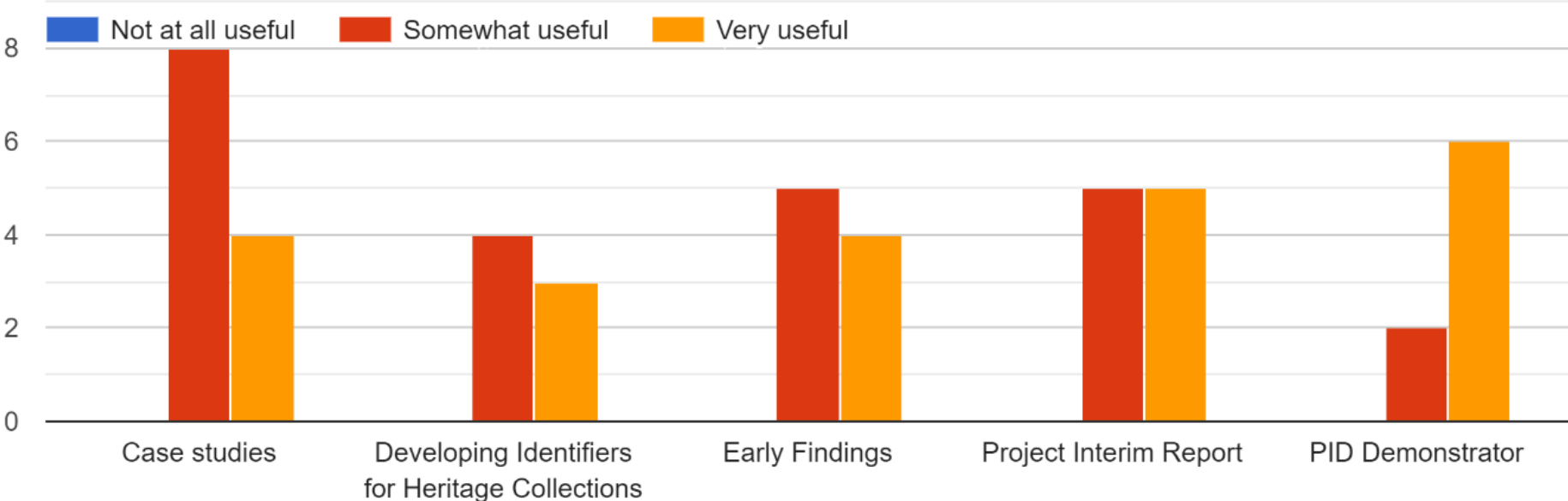


Figure 3: Responses to 'Did you find the output [selected in the previous question] useful?' from Survey 2. Non-UK responses are included.



Projects outputs

The table contains a list of project outputs.

Output	Description
Early Findings ¹⁹	This report outlines project activities to October 2020 and provides an overview of the awareness of PIDs across the sector at this stage in the project. It showcases the requirements of the sector and barriers to adoption we face. In turn, it aims to offer initial solutions to the challenges identified up to this point.
Persistent Identifiers at the British Library (Case Study) ¹⁹	This case study provides an overview of persistent identifiers at the British Library, including both those in use and those planned for the future. The aim is to help other heritage organisations to see what the path to use PIDs looks like, and understand what decisions need to be made along that path. Where possible we have tried to demonstrate ways in which organisations can engage and adopt PIDs in their processes and highlight considerations and challenges that may be encountered. This includes the lessons we have learned in facing those challenges, so that our peer organisations can avoid some of those pitfalls.
Persistent Identifiers at the National Gallery (Case Study) ²⁰	The National Gallery has conducted exploratory work related to persistent identifiers (PIDs) over the last number of years. This work has led to the development of a beta PID system based on URIs that is now transitioning to production. This case study provides an overview of the implementation and lessons learned through it. The National Gallery is home to the UK's national collection of paintings in the Western European tradition, comprising 2,300 paintings. It also houses additional material including a 'History Collection'; samples relating to paintings both within and external to the Gallery; a collection of digital images and frames; and library and archive collections.
Annotate It! Demonstrator ²¹	A bookmarklet demonstrator to illustrate how PIDs can be used to connect research with items in heritage collections.
Persistent Identifiers at the Natural History Museum (Case Study) ²²	This report describes the use of identifiers and PIDs at the Natural History Museum (NHM), London. The NHM is in the midst of an extensive collection digitisation programme to make all of the specimens in its collections available online, almost 4.8 million of the 80 million specimens are available so far. The NHM is creating a new data model to document complex digital objects more

¹⁹ Madden, F., & Kotarski, R. (2020). Persistent Identifiers at the British Library. British Library.

<https://doi.org/10.23636/1242>

²⁰ Madden, F., & Padfield, J. (2020). Persistent Identifiers at the National Gallery. British Library.

<https://doi.org/10.23636/1243>

²¹ Roderic Page. (2021). rdmpage/pid-demonstrator v.01.1 (v.01.1). Zenodo. <https://doi.org/10.5281/zenodo.4560194>

²² Madden, F., & Woodburn, M. (2021). Persistent Identifiers at the Natural History Museum. British Library.

<https://doi.org/10.22020/K99S-WE61>

	effectively, of which identifiers will form a core part. The NHM's Data Portal forms the main external point of access for the NHM's research and specimen collections. The digitised specimen collections, currently numbering 4.8 million, are assigned Globally Unique Identifiers (GUIDs) which form citable versioned links to records. The NHM mints Digital Object Identifiers (DOIs) for datasets created by staff and researchers affiliated with the Museum. As much of the data is tabular, the Data Portal allows for DOIs to be minted for each user query as needed, so their retrieved data can be cited and re-retrieved.
Developing Identifiers for Heritage Collections ²³	A guidance resource illustrating the different types of PID implementation and the requirements for each type. It also provides guidance on various aspects of implementing PIDs including how to guarantee persistence and the costs associated with implementing PIDs. This research and the project website were built using the Simple Site system.
Persistent Identifiers at the Royal Botanic Garden Edinburgh (Case Study) ²⁴	This case study provides an overview of the use of persistent identifiers at the Royal Botanic Garden Edinburgh (RBGE). RBGE delivers plant science, conservation and education programmes that are underpinned by the world-class RBGE Collections, comprising a Living Collection, a Herbarium collection and Library and Archive collections.
Video ²⁵	Persistent Identifiers help make collections available for the long term so they can be discovered, researched and cited. In this video, a range of experts describe how PIDs can help manage, research and digitise collections. Shorter edits illustrate how PIDs help particular roles in heritage organisations.
British Library Policy on Persistent Identifiers ²⁶	The British Library developed its Policy on Persistent Identifiers to act as an exemplar for peer organisations, but also in support of decision-making within the Library and on-going strategic infrastructure projects.

²³ Padfield, J. (2021). Simple Site - with Dynamic Build Option (v1.7). <https://doi.org/10.5281/zenodo.5137663> (Live version at <https://jpadfield.github.io/simple-site/>)

²⁴ Madden, F., & Mitchell, L. (2021). Persistent Identifiers at Royal Botanic Garden Edinburgh. British Library. <https://doi.org/10.23636/PVFS-N308>

²⁵ British Library. (2021). Persistent Identifiers in Cultural Heritage Collections. British Library. <https://doi.org/10.23636/13KA-JE14>

²⁶ British Library. (2021). British Library Persistent Identifier Policy. British Library. <https://doi.org/10.23636/KWGH-PC35>

Recommendations for the programme

Rec.#	Recommendation	Status	Source
R1	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	Completed	Early Findings
R2A R2B	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.	Completed	Early Findings
R3	We strongly recommend that heritage organisations start to work with their system suppliers to ensure systems meet their PID-based requirements and community principles	Adapted and On-going	Early Findings
R4	The TaNC programme should look at opportunities to continue to gather cost information on PID implementation from organisations just starting use of PIDs, in particular across a more diverse sample of organisations. Costs should consider professional development in support of better use of integrated PIDs, as well as stand-alone PID implementation costs	Adapted and On-going	Early Findings
R5	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	Completed	Early Findings
R6	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 began this work and has made further recommendations on a sector-wide approach, but these will need to be tested and refined by Towards a National Collection's Discovery Projects	Completed	Early Findings
R7	More IROs, higher education institutions and heritage organisations should implement policies on the use of PIDs to support linking of items and their metadata across institutional	On-going	Early Findings

	boundaries, and identify a minimum technical passive provision for PIDs that future-proofs new tools and systems for their use		
R8A R8B	Where key strategic systems cannot be easily reworked for PID use: A: lightweight add-on software 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.	On-going	Early Findings
R9	The TaNC programme, in collaboration with IROs, heritage organisations, higher education institutions and future projects, should define sector-wide principles for an approach to identifiers built on common principles, functionality, and use cases, such as those outlined within ‘Developing Identifiers’. This approach does not necessarily require all organisations to use the exact same identifier, and so enables the benefits of wider take-up, while respecting the diverse management needs, processes, and resource constraints of organisations, and avoiding use of ‘PIDs for PIDs sake’	Adapted and On-going	Early Findings
R10	The guidance within ‘Developing Identifiers’ should be used as a common starting point for the sector, allowing a common understanding and articulation of requirements	To commence	Final Report
R11	A path to community adoption of the resource will allow it to be maintained, updated, and adapted as use of persistent identifiers in the community evolves, and we call on the community to use, contribute to, and adapt the resource	To commence	Final Report

The table above summarises the project’s final recommendations, including completed activity from the project’s earlier findings, and new and adapted recommendations for the programme. In relation to recommendation 4, the approaches taken by case study organisations were such that cost profiles for that work were not possible to extract during the project. Some existing explorations of costs were pulled together²⁷ and can be read in conjunction with a recent report from UKRI¹⁰ that does outline the scope of similar PID activity costs. A first step will be to analyse how these existing costs map to the UK heritage sector. It was remarked especially clearly at the project’s final event, that wider access to technical infrastructure across the sector, particularly for those organisations with fewer resources, are a clear need, underlining Recommendation 8B. Persistent identifiers can easily be integrated into new tools for digital hosting of content online, considerably lowering the costs and expertise required for PID adoption compared to adoption as a discrete and separate activity. While bought-in vendor solutions may have appropriate PIDs

²⁷ Madden, F. (2021). How much does it cost to implement PIDs? [Blog post]. <https://tanc-ahrc.github.io/PIDResources/Cost%20of%20implementing%20identifiers.html>

built-in, staff still need appropriate training and guidance to ensure the PIDs are used to their full benefit. This shifts the costs from implementation technologies and activities, to professional development. As a result, Recommendation 4 is adapted to obtain better cost information from organisations earlier in their PID implementation journey, as well as considering both implementation paths.

In support of Recommendation 7, The British Library has published its own policy on use of persistent identifiers across the organisation. Inspired by the identifier policy of the Bibliothèque Nationale de France²⁸, it sets out the features of PIDs that it will aim to use in future, as well as the roles that the organisation may take in supporting or developing identifier schemes that meet its needs and requirements. The Library will report on how well the policy proceeds as we encourage bodies across the sector to develop their own - either using the Library's policy as an exemplar and a starting point, or one that directly addresses their own collection needs.

There is little evidence from the latter stages of this project that any changes are required for Recommendation 8. While the sector should start to ensure that new strategic infrastructure supports the creation of PIDs for content and use of PIDs in metadata, the light-weight approach will support initial steps until then. The lack of infrastructure and resource (both technical and staff-wise) for smaller organisations is still the biggest issue for their participation in connected digital collections. While discussion with vendors and a better analysis of costing of infrastructure implementation as a broader topic will continue to help define and start to reduce these issues, a somewhat centralised provision of digital infrastructure (which may include collection identifiers or resolution services, through to vocabularies and metadata identifiers) may be the only path for some organisations to participate.

Recommendation 9 is the key point for the TaNC programme to push forward. We have seen through the diversity of the approaches to persistent identifiers that already exist, that a sector-wide approach cannot and should not be overly prescriptive in the types of persistent identifiers that should be used. Individual organisational needs (e.g. linked data metadata approaches vs. collection identification vs. machine readability as outlined in our use case mapping²⁹) and capacity both vary considerably and determine the most appropriate identifier tool(s). There are clear sector-wide benefits to a networked approach, in terms of making connections between collections, enabling FAIR collections, metrics, and reducing management burden. But a networked approach can still be built on common principles, functionality, and use cases such as those outlined within the *'Developing Identifiers'* guidance, without necessarily requiring all organisations to use the exact same identifier tools. Recommendation 3 refers to community principles as a point of discussion with system vendors. An approach based on common principles can speak closely to the needs of the sector, to avoid approaches that implement 'PIDs for PIDs sake' and that deliver measurable benefits.

We do recommend that the framework of requirements produced by this project for the *'Developing Identifiers'* resource, should be used as a starting point for the sector, to allow a common understanding and articulation of those requirements. We note that new questions will arise and so the guidance will need to be expanded as conversations in the community continue to evolve. Recent examples from the last of our case studies highlighted the question of using externally governed 'persistent' identifiers versus internally developed 'stable identifiers'; and where internal identifiers are developed, how should a namespace be chosen for that? This leads us to recommend that a path to community adoption of the resource should be

²⁸ Bibliothèque Nationale de France. (2021) Politique identifiants de la Bibliothèque nationale de France : déclaration de principes https://www.bnf.fr/sites/default/files/2021-01/BnF_politique_identifiants_principes.pdf

²⁹ 'Which persistent identifier does what? PID functionalities' <https://tanc-ahrc.github.io/PIDResources/PID%20Functionalities.html>

found so that it cannot just be maintained, but also updated and adapted as use of persistent identifiers in the community widens.

Development of the guidance resource on Github was intentional - it can be copied by anyone who wishes to adapt the material for their community, but most importantly anyone can contribute requests to update the content. We welcome continued contributions to the resource as a community effort and call on colleagues to continue to feed into development of the resource either: as a user, by feeding in new requirements and questions (by raising Issues on Github or adding comments and questions on pidforum.org); as a contributor, by issuing pull requests to make improvements and additions to the toolkit; or as an adapter, who can fork and tailor the resource for their specific community.

Contacts

The key contact for this persistent identifier related work is Rachael Kotarski, rachael.kotarski@bl.uk

Annex

Comparison of survey responses

Demographics

While the second survey attracted slightly fewer responses, at least 42.6% of the respondents had not completed the first survey, a figure which may even be as high as 75% if those ‘not sure’ didn’t respond to the first survey (see Figure A1).

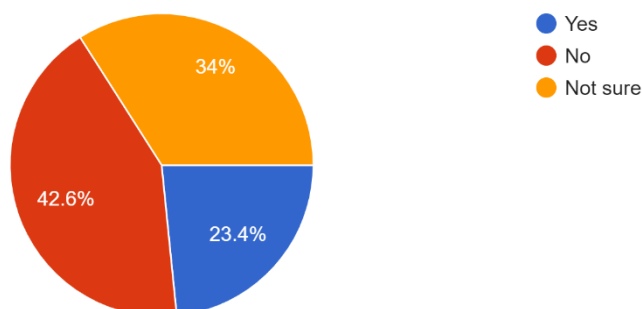


Figure A1: Responses to Survey 2: Did you complete the first survey of persistent identifiers in UK heritage? (n=47)

The role types of respondents were comparable, with Archivist, Research, Technology, and Collection Care or Management being the top responses across both surveys (See Figure A2). There was a notable increase in the numbers selecting Collection Care or Management in Survey 2, nearly doubling.

As for Survey 1, Survey 2 had good representation from outside the UK, with the majority of UK respondents being from England (see Figure A3). The following results exclude non-UK responses unless otherwise stated.

Organisational responses show a much better completion of both surveys although numbers are too low to draw quantitative comparison of any questions. Across 15 organisational responses to Survey 2, seven (46.7%) of these organisations also responded to Survey 1.

Figure A2: LEFT: Responses from Survey 1 showing the distribution of role types. RIGHT: Responses from Survey 2 showing the distribution of role types.

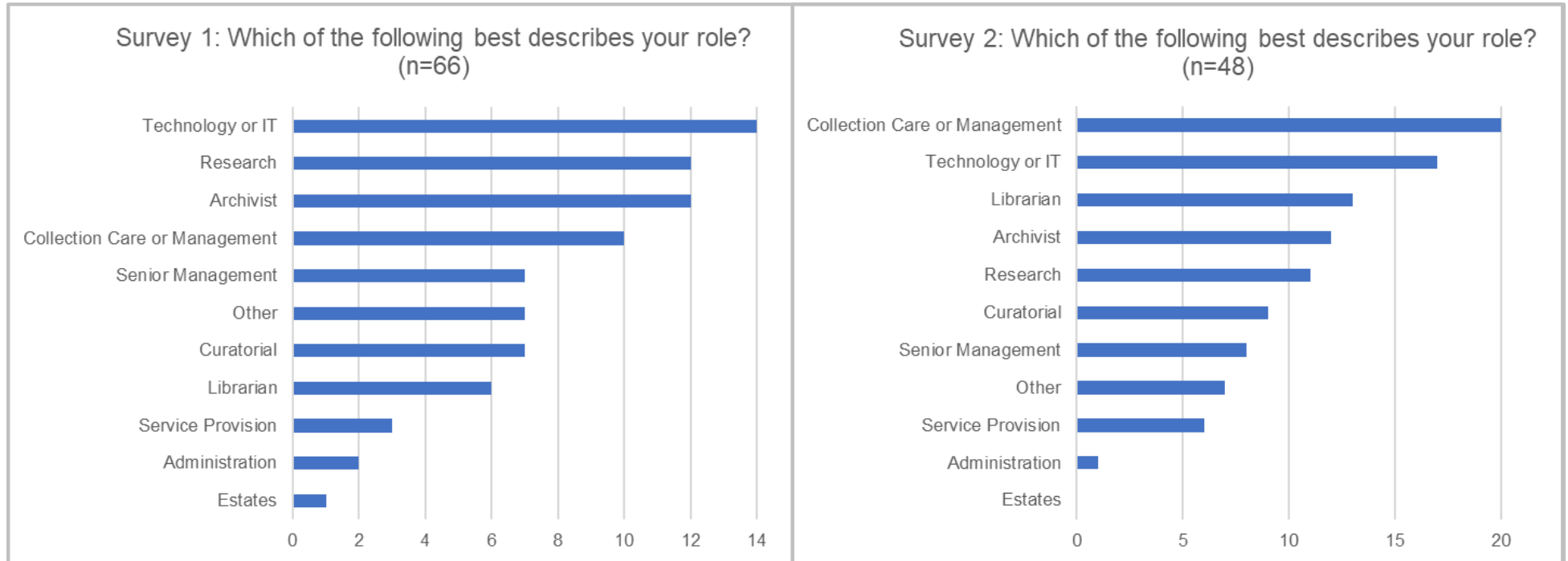
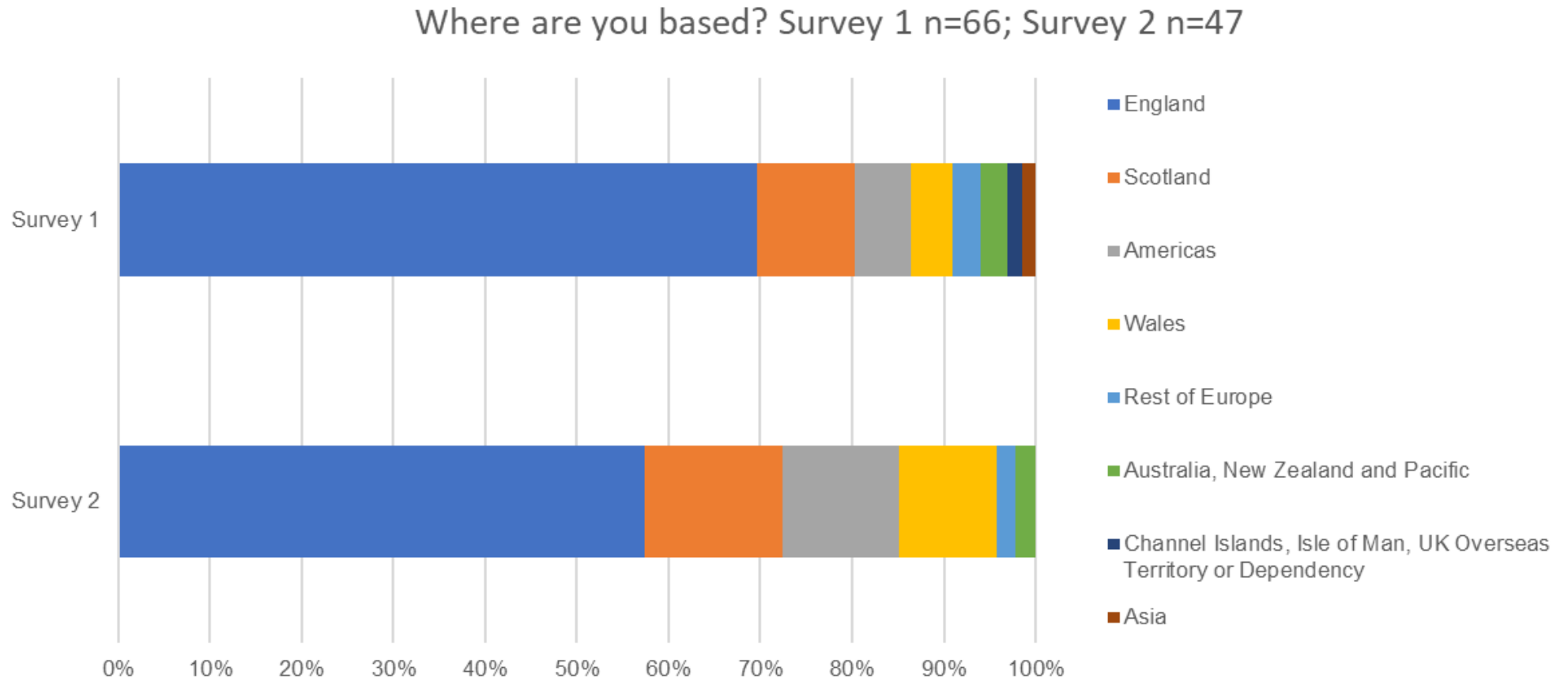


Figure A3: Locations of respondents from both surveys.



Awareness and use of PIDs

The results were broadly in keeping with the first set of results. Survey results suggest that there is awareness of PIDs as a technology, this hasn't increased in the samples from our surveys, although there is a slight increase in the use of some persistent identifiers over the period of the project (see Figure A4). At first glance this does not appear to be very successful, but 50% of respondents did indicate a positive change in their awareness (Figure A5).

In response to a question around what have you learned as a result of newly acquired knowledge of PIDs, the 21 free text answers received included specific activities in support of PIDs and their adoption such as starting *'an identifiers group within the organisation to agree how we want to manage them across the organisation'* and *'Develop policies to assist in management of development of PIDs'*. This demonstrates that the project has incited some activity in the sector, although there is still work to be done.

Looking towards organisational responses, there continues to be 'some awareness' across organisations, although numbers are too low for notable changes (Figure A6).

This increased familiarity expands to broaden the scope of the identifiers that respondents were aware of. Of the different types of identifiers, ORCID and DOI were the most familiar to respondents. Discipline specific identifiers such as IGSN and LSID were the least familiar, which matches well with results from Survey 1. The proportion of respondents using ARKs has increased between surveys; 47.5% had heard of it in Survey 1 but in Survey 2 this was 58.6%.

Several other PID types were mentioned, including Wikidata, VIAF, Grid, ISSN and ISBN. While these were the most commonly mentioned across both surveys, in Survey 1, CETAF Stable Identifiers were mentioned more frequently. Despite significant overlap, several identifier types were only mentioned in one survey or the other. In Survey 2, several geographical identifier schemes were mentioned, e.g. Ordnance Survey Identifier (OSID), Unique Property Reference Number (UPRN) as well as EAN and GRSCICOLL IDs, which are legacy IDs used in GBIF collections. IDs mentioned in Survey 1 include NSIDs, Gazetteer URIs and Europeana IDs. These disciplinary-specific identifiers were only mentioned once or twice in all cases, indicating a strong bias based on individual responses.

Figure A4: Responses to 'How familiar are you with persistent identifiers?'

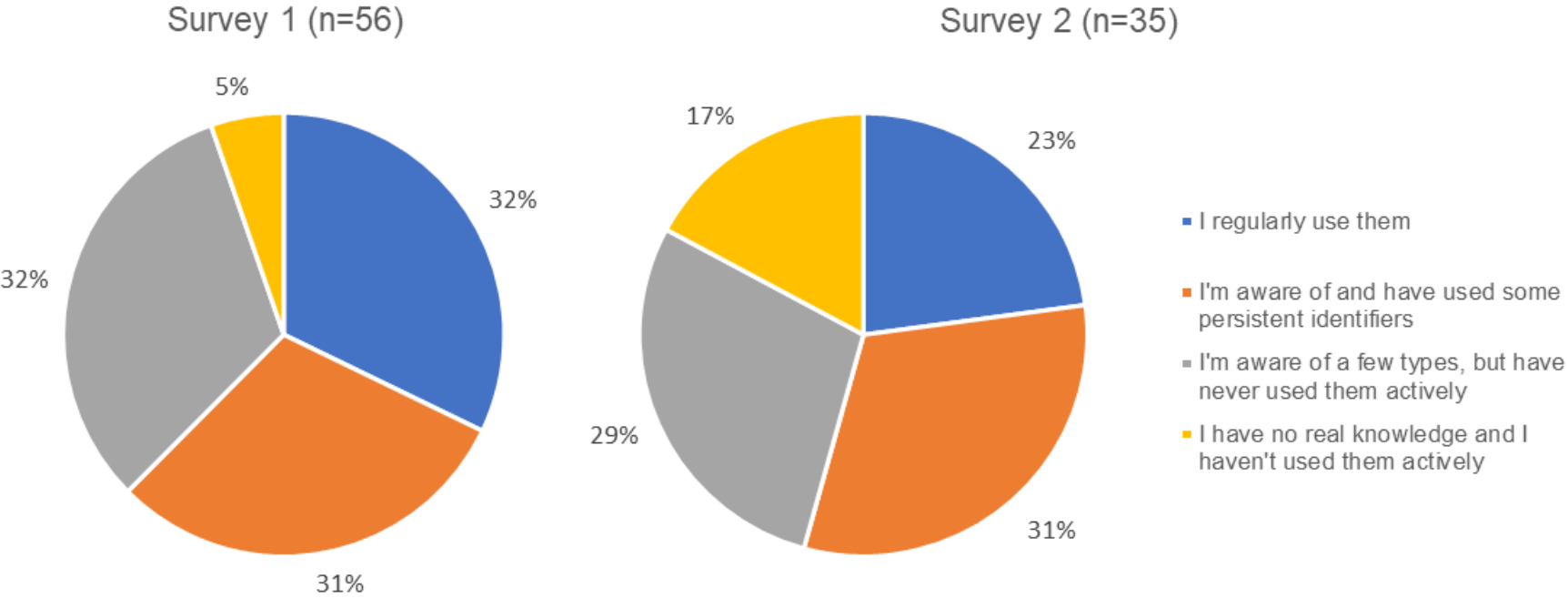


Figure A5: Respondents to Survey 2 indicate that there has been some increase in knowledge or awareness of PIDs over the duration of the project.

Survey 2 only: Has your awareness of persistent identifiers increased over the past 18 months? (n=35)

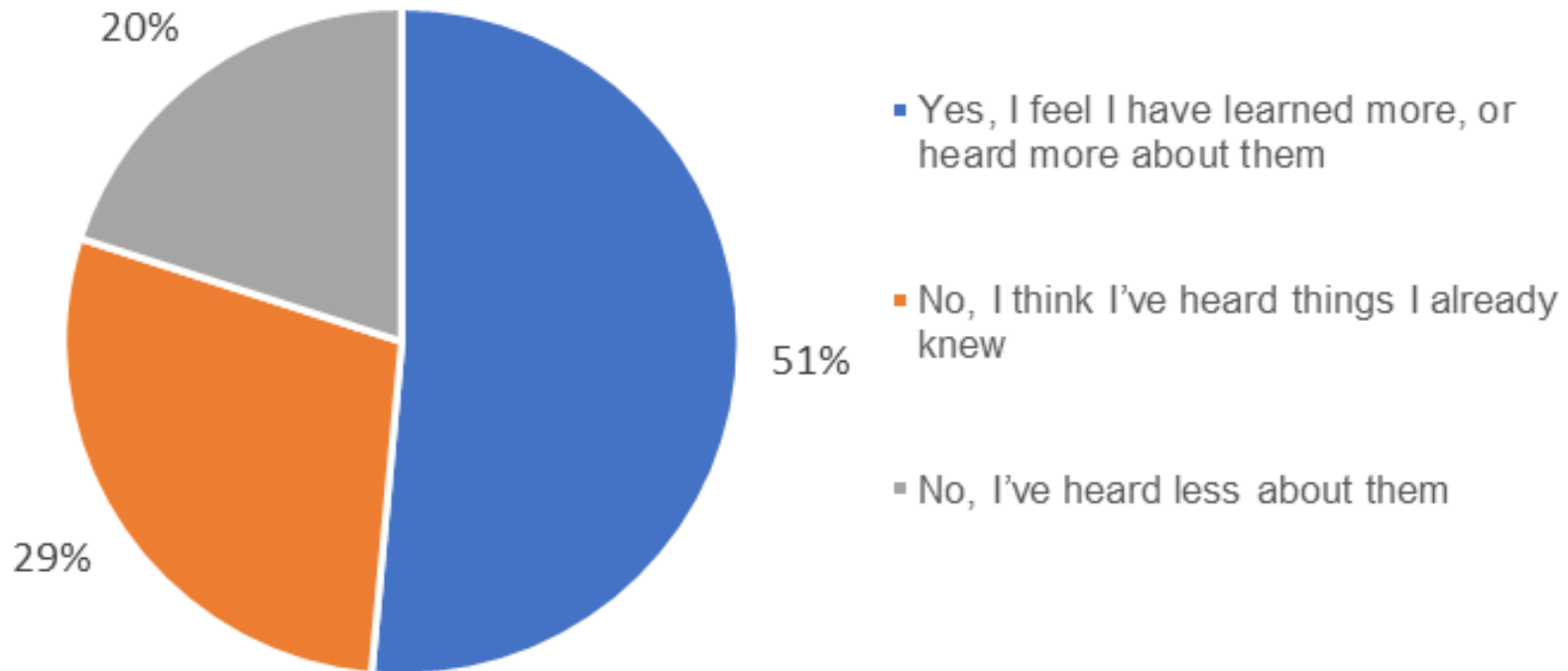
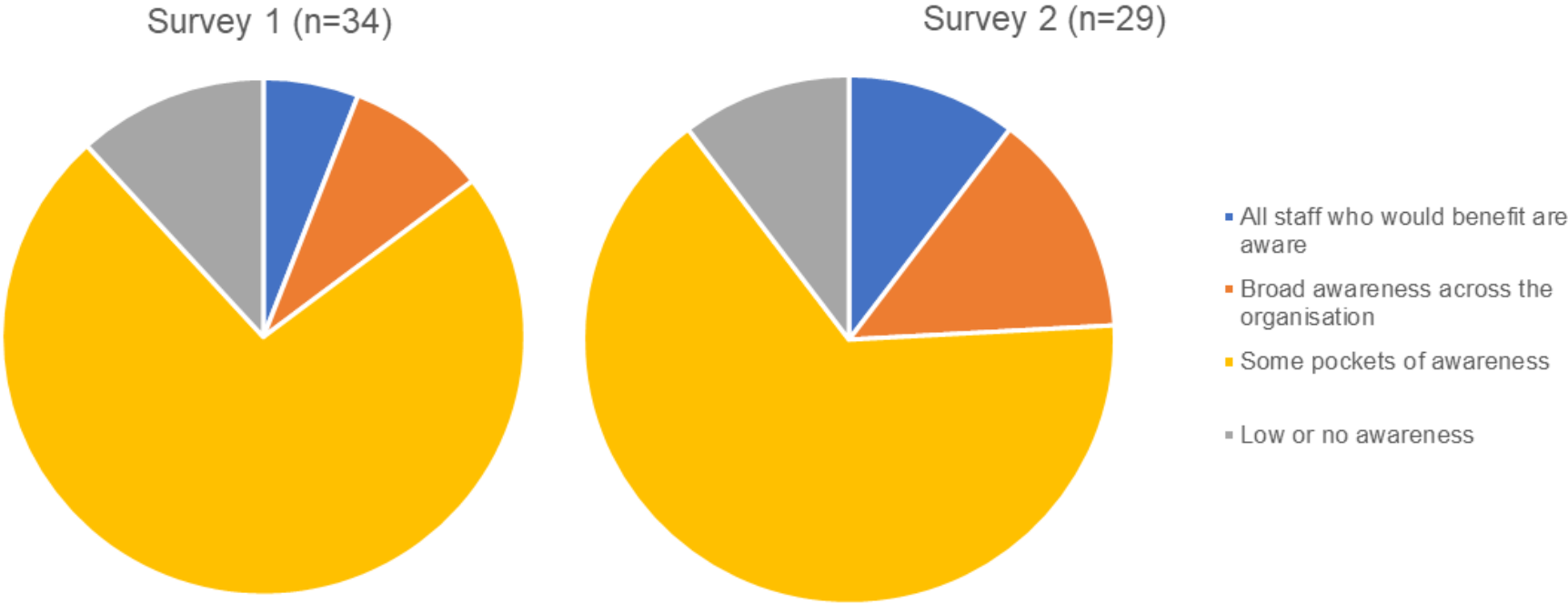


Figure A6: Responses to 'How much awareness of persistent identifiers do you think there is among staff across the organisation?'. Includes non-UK responses.



Barriers

Responses from Survey 2 indicate that awareness as a barrier to PID use in organisations and across the community is decreasing, thus increasing the importance of overcoming the barriers of resources and technical issues (see Figure A7).

The free text comments around barriers indicated that in certain institutions the fact that they will not raise additional income but do involve implementation costs is a blocker. Another said: *'There is a strong awareness of the need, it is just bringing the moving parts together and finding resource.'* Another personal reflection was *'that the benefit of having stable record identifiers is not sufficient that it warrants the effort to ensure long term stability. This results in it not getting the attention and commitment necessary to achieve it.'* Others mentioned competing projects and the scale of the data means resources can be limited.

These kinds of responses across both surveys show that despite the efforts of the project, the benefits of PIDs to the sector remain unclear for decision-makers. Some more nuanced views also emerged around a lack of PID creation and management within the technical solutions which are in use by organisations.

"At a senior level the organisation sees no benefit in using PIDs. Some professionals see potential benefits to the community and PIDs may therefore be introduced at a project level where it can be done with zero additional cost and zero IT staff input." - Free text response to Survey 2

Project Impact

In Survey 2 we asked what resources respondents had used and what they had found useful. 15 respondents had used the project's resources (Figure A8). Amongst respondents, the case studies have been the most heavily used outputs. However, the free text comments indicated that there was still demand for more outcomes, including case studies relating to smaller organisations. One response mentioned the resources were helpful in developing internal policy, especially the different types of identifiers described in Developing Identifiers for Heritage Collections.

"Thank you for all the promotion, dissemination and support materials over the course of the project."

"Now that we are aware of the resources this project has produced we will be looking [at] exploring them with great interest." - Free-text responses to Survey 2

Of those that had used project outputs, all found them somewhat, or very useful (Figure A9).

The final question of the survey asked what respondents would like to see in terms of continued support and community activity. Some of the responses included more funding across the sector to support work on identifiers and easier easy to access and use PIDs. Others requested an informal network or events to share knowledge and expertise. Others suggested that it would be useful to keep the project's resources updated. A few responses also mentioned the importance of system suppliers supporting PIDs, and their key role in ensuring adoption across smaller institutions especially. Other free text comments included the challenge of securing buy-in from senior management.

Figure A7: LEFT: Survey 1 reponses on barriers to PID adoption within respondents' organisation or community.
RIGHT: Responses to the same question from Survey 2.

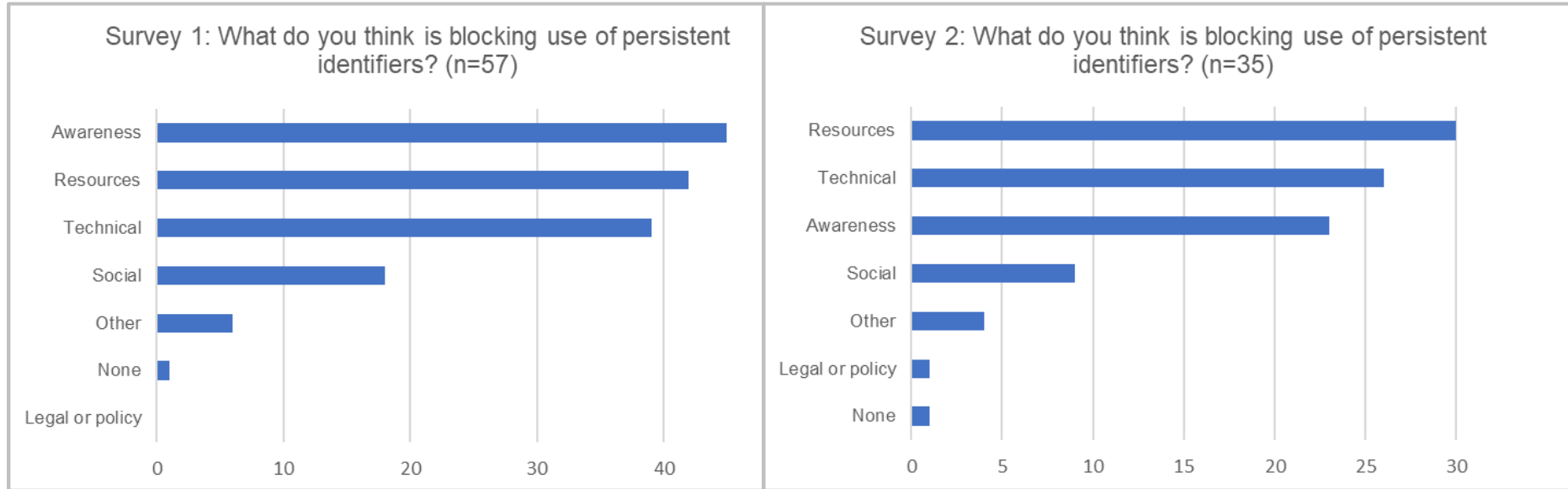


Figure A8: Responses from Survey 2 of which of the project outputs had been used. Non-UK responses are included.

Have you used any of the following outputs of the PIDs as IRO Infrastructure project? (n=48)

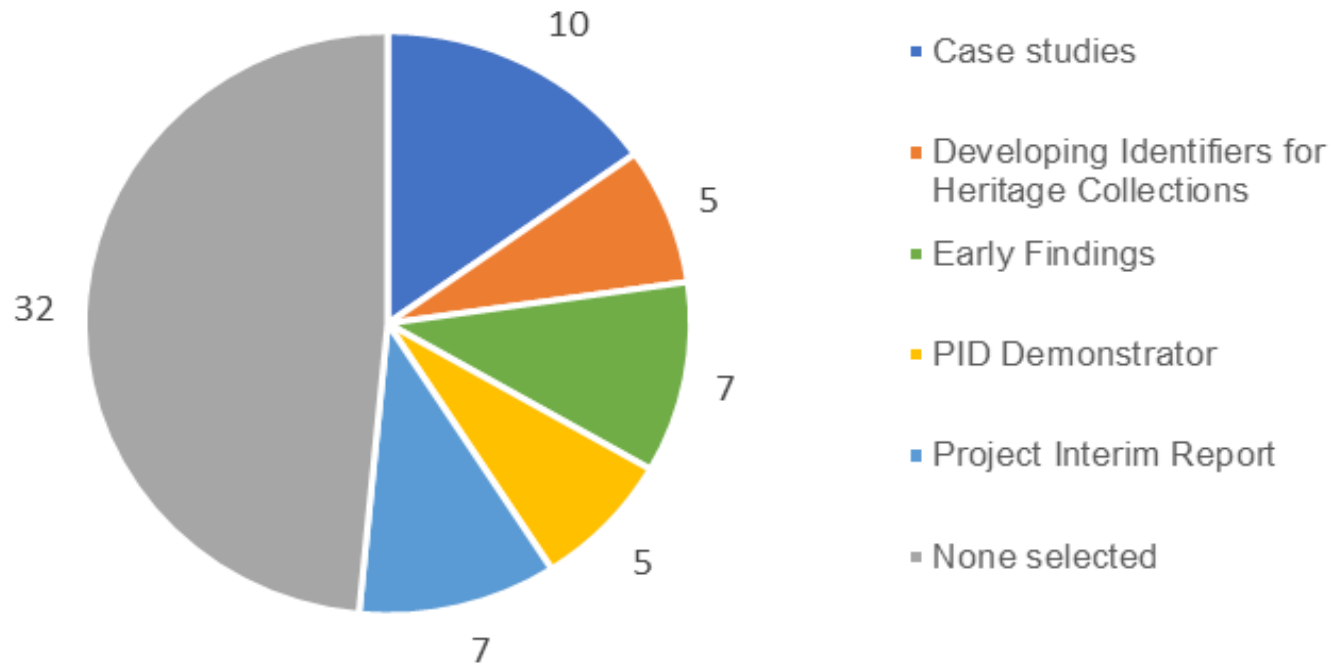


Figure A9: Responses to 'Did you find the output [selected in the previous question] useful?' from Survey 2. Non-UK responses are included.

