

Citation Capture

Enhancing Understanding of the Use of Unique and Distinct Collections within Academic Research and the Research Outputs Produced as a Result



Final Report

Delivered in partnership with Research Libraries UK,
The National Archives and Jisc

Table of Contents

1. Executive Summary	3
2. Research Overview	8
3. Model Citation Guidance	10
4. Embedding Best Practice	14
5. Data Sources and Tools	16
6. Implementation Partners	18
7. Conclusions	19
8. Appendices	22

Citation Capture © by Research Libraries UK, The National Archives and Jisc.

Citation Capture is licensed under a Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>.

This report was produced on behalf of the project partners by The Research Base.

www.theresearchbase.com | info@theresearchbase.com

1. Executive Summary

1.1 Project Overview

Research Libraries UK (RLUK), The National Archives (TNA) and Jisc are in the process of scoping the challenges and opportunities related to the development of standardised citation guidance appropriate for UK repositories that host unique and distinct collections (UDCs). While the project focuses primarily on the UK context, an awareness of international practices is also essential given that much academic research and associated publication happens across national borders.

Citation Capture: The Context

UDCs comprise original, primary sources across a range of media from cultural and literary heritage to audiovisual materials and born-digital content. UDC collections are housed or maintained online by a broad range of institutions, including research libraries, archives, cultural and heritage institutions, as well as online and digital institutions.¹

UDCs are widely used by academics, scholars and other researchers, with UDC materials frequently referenced within academic publications. Traditional academic references often take the form of an acronym formed from the letters of the repository where the UDC material is held. The increasing use of online platforms to publish and/or share research related to UDC materials has led to greater diversity in the style and format adopted for citations, including greater use of digital object identifiers (DOIs) and URL links.

Unique and Distinct Collections (UDCs)

'A collection that, regardless of format or location within an institution, derives significance from its interest to research, teaching or society through its association with a person, place or topic, such as to distinguish the constituent items from similar items which may exist elsewhere.'

– RLUK UDC Report (2014)

Academic and other scholarly publications are an important indicator of the use and value of UDC materials; however, at present, most UDC repositories have limited data on how frequently and in what format their materials are cited within both traditional academic publications and within online media. The benefits of improving citation capture for UDC repositories include:

- Providing metrics demonstrating the use, visibility and impact of UDC materials.
- Helping archivists, librarians and information managers make strategic decisions about how to develop their collections.
- Informing archivists, librarians and information managers about audience interest to help plan future events or projects.

There are a number of key challenges related to consistently capturing citation data for UDC repositories. These include inconsistencies in the way to which UDC repositories are referred within publications; a lack of awareness and technical expertise amongst archivists, librarians and information managers about how to capture citation data; and the need for a bespoke tool capable to capture and synthesise citation data from multiple sources.

The purpose of the current research project is to explore the nature and extent of these challenges and to identify future pathways to meet them. This is an ambitious project given the breadth of stakeholders that

¹ Cullingford, A. et. al. (2014). *Unique and Distinct Collections: Opportunities for Research Libraries*. London: RLUK.

would need to be engaged and 'buy-in' into the proposed model in order for it to become embedded across the archival, academic and publishing sectors. While the current project focuses primarily on the UK context, the project partners are also committed to developing a citation model and supporting guidance that could be replicated internationally, with the potential in the longer term for the model to come under the auspices of an appropriate international standards body, such as the International Association of Research Scholars and Administrators (IARSA).

Citation Capture Research Project

The foci of the current research were (i) to explore the scale of the challenge with respect to encouraging a greater degree of consistency in how UDC repositories are cited within academic publications and (ii) to identify key factors for successfully implementing standardised citation guidance for all UDC repositories.²

This report looks at the challenges and opportunities related to improving the consistency and accuracy of current citation practices for UDC repositories, as well as the strategies and tools required to effectively capture data for these citations. It is divided into four key strands: guidance, practice, data sources, and tools. It is our recommendation that these four strands, delivered together will prove to be the most effective approach to meet the long-term outcomes specified by the citation capture project, not least because it will allow project partners to approach the challenge at both user and systems level.

A final section exploring challenges and opportunities for implementation partners is also included within this report. A more detailed implementation plan has been developed to assess the actions and expected outcomes required to achieve the report recommendations; the implementation plan will also include a full risk register to support the further development and implementation of the citation capture project.

1.2 Key Findings

This report explores the potential to develop a standardised citation model that would improve the accuracy and consistency of citations to UDC repositories within academic publications. There were two key potential models identified during the primary and validation phases of research; we have included both models within this report and highlighted the benefit and challenges associated with each model in order to facilitate further discussion on the most suitable citation capture model going forwards.

In the primary research phase, key criteria for developing a standardised model identified by research participants (academics, UDC staff and project partners) included ease of integration with existing conventions; accessibility for academics and non-academics; and brevity of references.

Based on these criteria, evidence from this research phase indicated that a three-letter code using repository initials would be the most appropriate citation model. This model was deemed to offer the greatest affinity with current practice, while the potential drawback of the duplication of codes (where repositories share name initials, for example) could be mitigated within citation capture software as long as the full repository name is cited in the initial citation.

Participants at the validation workshop raised concerns about the suitability of the three-letter model, however, highlighting the need to create a model that would be suitable for a digital age, in which unique identifiers provide greater opportunity to track and monitor citation data. Participants also showed a high degree of confidence that younger, more technologically proficient repository staff and academics would easily adapt to a system that departed from existing conventions. As such, a model based on existing ARCHON codes was proposed that would provide unique identifiers that are digitally recognisable.

² <http://www.nationalarchives.gov.uk/help-with-your-research/discovery-help/find-an-archive>

Evidence from both research phases also indicates that a number of mutually reinforcing strategies may need to be adopted in order to achieve the overarching goal of the citation capture project, not only to embed best practice but also to gather more detailed and more accurate citation data. These strategies include:

- 🌀 Delivering a widespread publicity campaign that highlights the rationale behind the citation capture project and the benefits for both academics and UDC repositories of engaging with project outputs, and most importantly but not exclusively, the final citation model. Learned societies and leading research centres are likely to play an important role in any such publicity campaign.
- 🌀 Publishing comprehensive best practice guidelines on how to adopt the proposed model, as well as an up-to-date database of repository codes. There is also a need to encourage individual UDC repositories to provide citation guidelines; ideally with links back to the national best practice guidelines.
- 🌀 Providing UDC staff with the skills and training needed to develop their technical expertise. This is key to ensure that smaller repositories have the technical capacity to provide online user guidelines, as well as gathering and analysing citation capture data.
- 🌀 Encouraging the use of plug-ins across reference management software platforms, e.g. EndNote, to improve the accuracy of UDC repositories citations amongst the significant minority of academics that use these tools.
- 🌀 Garnering support from publishers. This is important given that publishing requirements are one of the key factors determining citation practices within academic publications.

There is also a strong case to be made for the project partners investing in developing bespoke citation capture tools that would automate the citation capture process. Bespoke software would ensure accurate and comprehensive data collections with users (such as UDC staff) only required to possess minimal technical expertise; another advantage would be the ability to respond to changes in how UDC collections are used, both now and in the future.

Existing data collection tools, such as Google Scholar and JSTOR for Researchers, offer an important starting point for the development of any bespoke citation capture tool. The principal drawback to these existing tools is that neither holds a comprehensive record of all potential UDC citations; as such, significant technical expertise is required to amass, clean and analyse the resulting data.

1.3 Recommendations

The following recommendations³ reflect the key areas for further development needed in order to achieve the long term goals of the citation capture project: (i) improve accuracy and consistency of UDC citations, and (ii) collate more reliable and effective data on citations by UDC repositories. These recommendations have been developed further in the implementation plan.

Recommendation One: Citation Model

The project partners will need to consider the relative benefits and challenges associated with the two potential citation models identified through this research project: namely, a three-letter code based on repository initials; and unique identifiers based on existing ARCHON codes.

The weighting of these benefits and challenges is likely to centre on the value attached to the key criteria that frame the development of a citation model. While integration with existing conventions and ease of use

³ Recommendations from the conclusions section of the main report are reproduced here in full because they are central to the reader's understanding should the Executive Summary be read as a standalone report.

were emphasised across the primary research, important considerations were raised during the validation workshop around ensuring that any citation model meets the needs of an increasingly technologically proficient workforce across both academic and archival sectors, as well as ease of use in relation to the growing number of both traditional archival resources available online and fully online repositories.

Recommendation Two: Centralised Guidance

The proposed citation model should be accompanied by centralised guidance, for example, through an online user-friendly webpage created and maintained by the project partners. This centralised guidance would be designed to provide repository staff and academics with information on how to implement the citation model; a database of codes that academics and other researchers can easily consult when creating citations to UDC repositories; and an overview of the broader citation capture project including its purpose and benefits for repositories, academics and other key stakeholders.

Recommendation Three: Assess Repository Capability

It is recommended that the project partners engage in a broader consultation process with UDC repositories in order to assess the extent to which staff possess the knowledge and skills required to improve citation capture practices. These include, for example, possessing the subject knowledge and technical expertise to upload citation guidelines to their websites, and using existing data sources to capture and analyse citations to UDC materials held by their repository.

This consultation process would focus on identifying clear strategies to successfully embed best practice within these institutions. One strategy to explore during this consultation process would be the setting up of a train-the-trainer model to support UDC representatives to develop the necessary skills and expertise to ensure best practice is embedded. Such training may include support to develop basic technical skills to run and maintain online best practice guidelines, catalogue plug-ins and citation capture practices.

Recommendation Four: Increase Knowledge and Awareness

The delivery of a widespread publicity campaign is recommended to increase awareness of the proposed citation model across all potential stakeholder groups, including academics and other researchers, publishers, learned societies and other disciplinary or professional associations. This campaign should focus on informing repositories and their users about the proposed citation model and the project's broader goals. Increasing knowledge and awareness is likely to go hand-in-hand with developing centralised guidance (Recommendation Two) so that individuals have ongoing access to relevant information about the citation model and broader citation capture project.

Delivery of this recommendation would also benefit from more widespread publication of best practice guidelines by UDC repositories, as this in turn may raise awareness of the proposed citation model, as well as encouraging more users — especially academics — to identify repositories as a key source of information on how to cite UDC materials, alongside existing disciplinary norms and publisher guidelines.

Recommendation Five: Improve Software Solutions

We recommend working closely with existing software companies in order to develop a streamlined approach to the process of capturing and analysing citation data. This may include, but not be limited to, initial discussions with ORCID with respect to creating unique digital identifiers for institutions based on their existing unique digital identifiers for individual researchers, as well as approaching software companies such as TurnItIn that hold a comprehensive dataset to explore the potential for using that dataset to capture citation data.

In the longer term, we also recommend that the project partners explore the possibility of developing bespoke citation capture software in order to achieve the long-term goals of the project. As well as overcoming challenges that users have experienced when using existing citation capture tools, this would create a valuable resource for UDC repositories as well as academics and other researchers.

If delivered alongside a significant publicity campaign to support the implementation of a standardised model (Recommendation Four), the development of improved citation capture tools is likely to make the benefits of following the proposed standardised model more tangible for repositories and their users.

Recommendation Six: Partnership Working

It is recommended that a consultation process is undertaken to assess levels of engagement among learned societies in terms of both the design and publication of the proposed model.

There is a strong likelihood that the project partners could successfully engage with learned societies in key disciplinary areas for UDC materials, given that many of them have a high number of academics in senior positions and that this stakeholder group has been highly engaged in the current phase of consultation. Project partners may wish to engage in a process of stakeholder mapping to identify the most relevant learned societies and, subsequently, to undertake a consultation process with these societies to assess levels of engagement in terms of both the design and publication of the proposed model.

Given the lack of engagement from publishers thus far, project partners may want to consider exploring (i) potential strategies to engage with publishers, and (ii) the feasibility of implementing the standardised model without widespread support from the publishing sector if this engagement process remains unsuccessful. The prospect of limited engagement from publishers further underscores the importance of developing a standardised model for citations that reflects current practices, as this increases the likelihood that academics and other researchers who implement the new model would be able to do so within existing requirements specified by publishers and editors.

2. Research Overview

2.1 Research Design

As noted in the project overview presented above, the current research project has been designed to support the project partners (RLUK, TNA and Jisc) to:

- Identify existing practices and/or examples of best practice with respect to UDC repository citations.
- Explore opportunities and challenges related to the development of a standardised model of citation for UDC repositories.
- Develop model citation guidance for UDC repositories.
- Create an implementation plan to promote uptake of the model citation guidance.

The project methodology included the following research elements: a review of relevant internal data and documentation; a light-touch literature review; qualitative interviews with project partners, academics and UDC representatives; a survey distributed to academics, UDC representatives and publishers; and a validation workshop held with senior archivists.

2.2 Methodology

Literature and Desk Review

A light-touch literature review was undertaken in order to identify key challenges in developing a model citation system for UDC repositories and highlight existing best practice. This included assessing the current landscape of UDC repositories in the UK; establishing current citation practices in relation to style/format used for UDC citations; reviewing international practice on UDC citation styles and guidance; identifying examples and/or case studies of best practice; and analysis trends for UDC citations across key social media platforms.

A desk review of relevant internal data and documentation was also completed to develop a deeper understanding of the project aims and objectives.

Interviews

Interviews were held with 29 academics, specialists working in UDCs and other experts working in associated fields in order to address the key questions identified in the research framework, and to explore relevant issues raised in the literature review. Interviews were delivered from 5th March to the 24th April 2018; all interviews were conducted over the phone.

The breakdown for interviews by participant type is summarised in the table below.

Group	Number
UDC Representatives	18
Academics and other users	8

Group	Number
Publishers	1
Project partners	2
Total	29

In addition to the completed interviews, we also received email feedback from one ‘research funder’ following a request for interview. This feedback has been incorporated into the full interview analysis in the appendix.

Survey

An online survey was disseminated to identify existing practice with respect to creating, managing and tracking UDC citations. The survey was designed by The Research Base with the project partners distributing the survey across their mailing lists. The survey included separate questions for UDC representatives, academics and other researchers, and publishers and funders. A total of 184 responses were received from participants, comprising 90 repository staff and 94 academics/researchers. Survey responses were received from 3 April to 27 April 2018.

Where survey questions were the same for repository staff and academics/researchers, responses were combined and analysed using SPSS. Responses from the different groups were then tested for statistical significance using independent samples t tests.

Validation Workshop

A validation workshop was held on July 27th 2018 with seven senior archivists and librarians based at repositories attached to academic institutions. Participants were invited to take part by the project partners as specialists within their sector; however, individual responses have been fully anonymised in all reporting to the project partners.

The purpose of the validation workshop was to test emerging findings from the primary research phase prior to finalising the full project report. Where feedback from the validation workshop differs from the primary phase of the research, this has been clearly indicated in the report below.

Analysis and Reporting

Evidence has been collated and analysed across the key research areas in order to identify emerging trends, as well as points of divergence, across key stakeholder groups. An initial phase of analysis was conducted in order to develop the model citation guidance⁴ with further analysis presented in the current citation capture report; a full summary of the evidence used to develop this report can also be consulted directly in the report’s appendices.⁵

⁴ Previously submitted as a separate document and summarised for reference below.

⁵ Report appendices include the full desk review, alongside interview and survey analysis relevant to the development of the model citation guidance.

2.3 Research Limitations

- There was limited engagement from publishers and funders with this research. Only one interview was held with a publisher with email feedback received from one funder; no survey responses were received from publishers or funders. While more research is needed to better understand publishers' and funders' perceptions of the benefits of greater standardisation in citation to UDC repositories, it should also be noted that funders face particular constraints in terms of taking a top-down approach to setting out specific requirements related to research delivery.
- It has not been possible to assess how existing citation practices differ across small, medium and large UDC repositories, as responses from survey participants regarding the size of their repository holdings or their repository's head count were not in a standardised format which would allow for a meaningful comparison. While some participants specified the size of their repository holdings in cubic metres, many used linear metres instead. For staff head count, some provided a figure for the full-time equivalent instead of the total head count, while only some included the number of volunteers. Some survey participants also appeared to have included figures for the entire institution, as opposed to figures that were specific to archive and special collections.

3. Model Citation Guidance

3.1 Overview

The initial research project explored the possibility of creating a standardised citation model for UDC repositories that would specify alphanumeric codes for each UDC repositories. The following section provides a summary of the criteria that would need to inform any such model; two possible models; and a brief assessment of the feasibility of these models.

The scope for the citation model was originally intended to cover both traditional academic publications and non-traditional outputs including citations to UDC repositories within online publications and social media posts. However, evidence from the literature review and primary research with academics indicates that there is a trend towards using hyperlinks rather than traditional citations across online and social media platforms. As such, the core recommendations for developing a citation model refer principally to traditional publications with the recognition that citation style and format is likely to differ for online and social media platforms (and that hyperlinks are perceived as preferable to traditional citations on these platforms).

3.2 Criteria for Citation Model

The following criteria were identified by research participants in the primary research phase as essential for the development of a standardised citation model for UDC repositories:

- Integration with existing conventions:** Both academics and UDC staff emphasised the need for the proposed citation model to be easily integrated into existing citation practices. For academics, this also included the flexibility to adapt the proposed citation model to existing citation styles, such as Harvard and Chicago, in order to meet publishers' requirements and to reflect disciplinary norms.
- Intuitive for both academic and non-academic users:** Academics highlighted the need for citations to be sufficiently detailed while also being easy for readers to understand and follow up, should they wish to consult the same sources. UDC staff also noted that the proposed citation model should be sufficiently straightforward, in order that non-academic researchers can understand and use it.

 **Brevity of references:** It is important that the proposed citation model allows citations to be produced with relative brevity, given that many academics publish articles and books subject to specific word limits.

Given that no single model can fully meet all criteria simultaneously, the scholarly community may need to explore further the relative weighting and value attached to each of these criteria.

3.3 Developing the Citation Model

As part of this research project, we explored a number of possible models based on the criteria specified above: namely, a three-letter repository code; a four (or more) letter repository code; and using existing ARCHON codes.

Analysis of the benefits and drawbacks of each model based on the primary research indicated that the three-letter model most closely aligns with criteria set out above, particularly integration with existing conventions. However, participants at the validation workshop emphasised the challenge associated with the three-letter model and expressed a strong preference for a model based around ARCHON codes instead.

The benefits and drawbacks of both the three-letter and ARCHON models are set out below. We have included both models within this report and highlighted the benefit and challenges associated with each model in order to facilitate further discussion on the most suitable citation capture model going forwards.

The implementation of a standardised citation model offers a valuable starting point to improve the accuracy and consistency of current citation practices. However, it is likely that further strategies around guidance, data sources and tools will be required to sit alongside the proposed model in order to achieve the project's long-term goal of improving the quality and breadth of data on citations to UDC repositories.

Proposed Citation Model: Three-Letter Codes

Three-Letter Codes

Feedback from academics and UDC representatives indicates that UDC repositories are most frequently cited according to the following conventions:

- Full repository name for the initial citation, e.g. The National Archives.
- Abbreviated letter code for subsequent citations based on the repository name, e.g. TNA.

Three-letter repository codes were the most commonly used among the academics surveyed (42% of participants), although other combinations of letter codes also appear to be used frequently. Three-letter codes were also most frequently referred to as the dominant trend by both academic and UDC staff interviewees, as well as being the most widely recommended practice based on the analysis of citation guidelines presented on UDC websites.

The advantages of designing a model that accords with the dominant trend for three-letter codes is that it is more likely to integrate fluidly with current practice and easily adapted to meet different referencing requirements; that it is intuitive for both researchers and readers to understand; and that it achieves relative brevity.

The main objective of this model would be to encourage greater consistency from all academics and researchers, rather than introducing a new system. As such, citation practices are likely to remain the same for many academics already well-versed in creating accurate and readable UDC citations.

This model presents a number of significant challenges when rolled out nationally, however, and even more so if rolled out internationally. At present, most researchers develop abbreviated letter codes ad hoc for each individual repository, typically based around repository initials. Existing trends in the use of letter codes may also be informed by common practice adopted by academics working in the same discipline and/or repository preference; 43% of the UDC representatives surveyed said there was an established and recognised abbreviation for their repository.

Standardising this system for all UDC repositories in the UK would lead to the duplication of certain three-letter codes, e.g. it is likely that at least four repositories would adopt the code 'BRO': Bath, Berkshire, Berwick-upon-Tweed and Bristol Records Offices. Among the UDC staff surveyed who said there was an established and recognised abbreviation for their repository, 21% believed that their abbreviation was not unique to their institution and 18% were unsure. As long as researchers consistently cite the full repository name in the initial citation, the duplication of three-letter codes would not necessarily be a barrier to citation capture across academic publications; however, the success of this model would depend on researchers consistently and accurately citing the full name of the repository in the first instance, as well as developing citation capture software sophisticated enough to match repository names and codes.

The use of duplicate three-letter codes is preferable to assigning unique three-letter codes to repositories, given that unique three-letter codes would be likely to require random combinations of letters for some repositories, which would not be intuitive. This would lead to both researchers and readers needing to consult a database of codes, a scenario that may inhibit buy-in across stakeholder groups without significant efforts to publicise and support uptake of randomised codes.

Three vs. Four (or more) Letter Codes

The use of four (or more) letter codes was also considered during the primary research phase because of the potential for each repository to be assigned a unique abbreviated letter code that would be intuitive for researchers and readers to identify and use; this may be especially important as the number of online repositories continues to grow. The use of four (or more) letters codes would also reduce the likelihood of duplication of abbreviated letter codes used by repositories internationally, many of which use three-letter codes for standardised abbreviations, e.g. The National Archives Australia (NAA).

The principal drawback that led to the three-letter model being considered preferable is that the use of four (or more) letters is less common than three-letter codes within current practice; feedback from both academics and UDC representatives suggests that most academics have ingrained habits when it comes to citations and that the most feasible model is likely to be that which requires changing the habits of the smallest proportion of academics possible. The adoption of four letter codes may also risk creating a dual system of citations meaning that citation capture practices would need to be tailored to identify both pre- and post-model citations.

Validation Workshop Feedback

Validation workshop participants identified a number of challenges associated with the proposed three-letter model, all of which centred on the idea that any proposed model needs to be suitable not just for the short term (i.e. reflecting current practices) but also able to adapt to the needs of future generations of academics accessing and referencing digital UDC resources and developing their own digital identities as researchers.

Specific concerns raised include (i) the belief that three-letter codes were not, in fact, currently common practice; (ii) a potential issue where two identical abbreviations could be used in the same paper, e.g. Capetown University Library and Cambridge University Library, with no means of distinguishing between them after the initial reference in full; and (iii) a sense that three-letter codes would not contribute to developing a system, but simply to making existing practice more consistent.

There was also a sense that the three-letter approach more closely aligns with existing practices of older academics (who will be leaving the profession sooner), whereas younger and mid-career researchers are more likely to adopt new practices as needed.

Alternative Citation Model: ARCHON Codes

There is contrasting evidence from the primary research and the validation workshop as to the suitability of ARCHON codes for a standardised citation model. ARCHON codes are existing unique codes to identify UK and a small number of international archives listed in the UK National Register of Archives. The current databases of ARCHON codes is maintained by The National Archives.⁶

Evidence from the primary research indicated that, while ARCHON codes offer an alternative pre-existing model of unique identifiers for UDC repositories, these numeric codes are not intuitive and would require users to look up codes for each citation. Primary research also indicated there is limited or no uptake of ARCHON codes within current practices for both academics and UDC representatives. As such, a citation model based on ARCHON codes was considered to depart significantly from the key criteria set out above.

However, feedback from the validation workshop emphasised the potential benefits of developing a citation model based on existing ARCHON codes, especially when compared to the proposed three-letter model. Benefits of ARCHON codes identified by participants include the following:

-  They are fixed/long-lasting and deal with any name changes which might occur.
-  They comply with international practice.
-  They are based in an established system.

A further advantage of developing unique identifiers for each repository is the potential to capture number of citations within each individual publication, as well as being usable across online and social media platforms without requiring the full repository name to be cited. The key to the success of using ARCHON codes, according to participants, centres on integrating these codes within existing practices (even if the codes themselves depart from current citation conventions). This would be likely to involve integrating ARCHON codes into catalogue references so that these codes are natural captured within citations: *'Once you start putting ARCHON codes in catalogues, [academics] will think it is part of the reference and practice will change.'*⁷

A challenge with the use of ARCHON codes is that many unique and printed collections do not currently have ARCHON codes attached and would need some other form of identifier. One participant mentioned ORCID codes as a platform for developing codes for these resources,⁸ with the recognition that any system accompanying ARCHON, and facilitating the inclusion of repositories without an ARCHON code, needs to be consistent and unique.

3.4 Opportunities and Challenges

In addition to the opportunities and challenges identified above that are associated with each model, there are a number of general considerations that apply to both:

⁶ <http://discovery.nationalarchives.gov.uk/find-an-archive>

⁷ Validation workshop.

⁸ <https://orcid.org/organizations/research-orgs/resources>

- 🌀 **Repository name changes:** Repositories are often subject to mergers and name changes; any citation model that uses abbreviated letter-codes based on repository initials would need sufficient flexibility to accommodate these changes, and to adapt citation capture practices accordingly.
- 🌀 **Online and social media citations:** Citations on online and social media platforms typically use hyperlinks. Capturing citations on these platforms is likely to be dependent on analysis of the domain names attached to repositories in order to build up a comprehensive picture of citations, in addition to the proposed citation model/s for traditional academic publications laid out above.
- 🌀 **Reference management software:** Encouraging researchers to adopt any proposed citation model and to apply that model accurately is likely to be the main challenge in terms of generating citations with sufficient accuracy and detail to allow for improved citation capture. This challenge may be partly mitigated by working closely with reference management software companies such as EndNote to develop plug-ins to improve the accuracy of citations amongst users of these software tools.
- 🌀 **Single point of access for citation guidance:** Both academics and UDC representatives referred to the benefits of creating centralised citation guidance to specify how the proposed model would work in practice, as well as providing a database of codes to cover all UDC repositories included within the model. Validation workshop participants also emphasised the benefits of developing centralised guidance, while noting that if ARCHON codes are selected as the citation model, a crucial element will be explaining what ARCHON codes are, as well as having breakdowns of references so that it is clear how the ARCHON code fits within a standardised citation.

Overall, it is clear from both the primary research and validation workshop findings that there remain potential challenges attached to identifying a citation model that meets the needs of both the archival and academic sectors. The crux of this debate appears to focus on the different degrees of emphasis placed by sector representatives on the criteria required to develop a successful standardised citation model; in other words, the choice between a system that integrates closely with existing practice creating minimal burden on users vs. developing a system of unique identifiers suitable for the digital era.

4. Embedding Best Practice

4.1 Overview

There are a number of emerging trends from the evidence that point towards the need to invest time and resources in ensuring that best practice becomes embedded across key stakeholder groups.

Evidence from both survey and interview participants highlights that relatively few repositories currently offer online citation guidance to users. Almost all (90%) of the UDC representatives surveyed who said they provide citation guidance for their users do so via email or on request, while less than a third (27%) provide guidance on the repository webpage. Likewise, over two thirds of interviewees stated that they provide guidance on request but only one third referred to having online guidance in place.

Academics also report that they rarely use repository guidelines to develop citations to UDC materials, although this may be partly tied to the current lack of guidance provided by repositories. Interviewees also highlighted the need for a significant publicity campaign targeting academics and other researchers that access and cite UDC materials in order to raise awareness.

Practices related to citation capture across UDC repositories also appear to be under-developed. More than half (57%) of the UDC representatives surveyed said they did not collect any data on published citations to their collections. Even among UDC repositories that already capture citations, this appears to be done mostly on an ad hoc basis through word of mouth and/or generalised searches rather than being

systematically captured using existing tools and software. A similar trend was also identified across interviews with UDC staff, with interviewees explaining that low levels of technical expertise and lack of awareness were the principal barriers to capturing citation data. Except for Google Scholar, UDC representatives did not appear to be very familiar with existing citation capture tools.

4.2 Key Strategies

Embedding best practice is likely to involve a significant promotional campaign, as well as the development of best practice tools and guidelines, in order to ensure that any proposed model achieves buy-in across stakeholder groups. Key strategies to embed best practice may include the following activities.

Best Practice Guidelines

Publicising the proposed standardised model is essential to ensure user uptake. Repositories have the potential to play a key role in making sure that their users are aware of a standardised style. The National Archives could take a leading role in encouraging UDC repositories to make best practice guidelines available to all users, both on the repository webpage and clearly displayed within the repository.

Repositories should also be encouraged to provide researchers with a link to the proposed centralised guidance discussed in Section 3.4 within their own institutional guidelines, in order to raise awareness of the sector-wide approach to UDC citations. Workshop participants also suggested repository staff could include information about codes within email footers to publicise individual repository codes.

Publicity Campaign

There is strong support from both the primary research and validation workshop for a widespread and ongoing publicity campaign. The purpose of the campaign would be to ensure that UDC repositories, academics and other non-academic researchers are aware of the proposed citation model, including its benefits for the archival, library and scholarly community, as well as signposting the centralised guidelines setting out how to produce citations in accordance with the proposed model.

Participants at the validation workshop highlighted additional benefits for academics associated with an ARCHON-based model that could be communicated within a publicity campaign. These include the potential to use ORCID codes, in conjunction with ARCHON, to link academics with others who are accessing the same resources, as well as better capture of academics' own publications, and for individual repositories to be able to link to their articles. Workshop participants also noted that training on the citation model could be integrated into existing early career researcher training to reach the incoming generation of academics.

Citation Capture Practices

UDC repositories will likely need a considerable degree of encouragement and technical support in order to capitalise fully on the putative benefits of a standardised model. This is likely to include both raising awareness and developing technical skills, especially for smaller repositories that may not have access to technical specialists in-house.

In the short to medium term, before any finalised citation model is fully taken up, training should be the priority in terms of supporting staff to capture and analyse citation data. This could be done, for example, through half a dozen regional workshops involving a cascade training model, plus online workshops, and perhaps a MOOC. Where possible, existing means of training should be utilised; for example, training could also be filtered through Archives and Records Association (ARA) and Archives Hub, and larger repositories could help to push out any centralised information to smaller ones.

Catalogue Plug-ins

There is a strong case for encouraging the uptake of catalogue plug-ins that allow repository users to export UDC reference data directly from repository catalogues into their preferred reference management software (e.g. EndNote). This would increase the accuracy and consistency of UDC citations. However, only half of academics interviewed referred to using this kind of software and almost three quarters of academics surveyed stated that they did not use any such software, although evidence from interviews suggests that usage trends are partly generational with older academics less likely to make use of reference management software. Technical support to integrate plug-ins into existing catalogue systems would also need to be provided by repositories' systems suppliers as there are low levels of technical expertise reported within many smaller repositories.

As noted in Section 3.3 above, workshop participants also emphasised the benefits of integrating repository codes attached to the citation model within each individual catalogue reference itself. Both CALM and ATOM software were referred to as the key software platforms to achieve this with ATOM seen as particularly beneficial given that it offers opportunities to develop open source software to meet the needs of smaller repositories.

Publisher Support

Support from publishers for a standardised citation model would have a considerable impact in terms of embedding best practice. Publishers' guidelines were viewed to be the most important factor influencing the choice of citation model among the academics surveyed, followed by disciplinary norms and personal preferences. At present, UDC repository guidelines do not appear to be considered very influential although this may change if repositories offered more detailed and consistent guidelines for their users. Funder requirements were considered the least important within the primary research, which suggests that funders may not need to take a prominent role in ensuring uptake for a standardised model.

5. Data Sources and Tools

5.1 Overview

As noted above, the success of the citation capture project relies upon a number of mutually reinforcing strands of activity to include improving the technological infrastructure so that more accurate and reliable citation data can be captured. Two clear strands that could be pursued in order to improve citation capture through digital tools and software are: (i) increasing awareness and use of existing data sources and (ii) developing a bespoke citation capture tool that would automate the citation capture process.

Improving the use of existing data sources and/or developing bespoke tools would address a number of challenges associated with achieving the broader goals of the citation capture project. In short, there is a risk that any model or approach to standardise citations would not be able to keep up with changes in how UDC collections are accessed, used and shared by users as a result of a growing mass of digital content and social media references. The growth in accessing UDC sources online also means that the ability of repository staff to guide and support their users in how to implement a standardised model is likely to be increasingly limited. Furthermore, the level of consistency among users in adhering to a standardised citation model may be limited as UDC repositories attempt to appeal to wider audiences; these audiences may be less receptive to a standardised approach than current user groups.

Investing in better citation capture tools would address the challenge of citation capture at a systems level rather than at an institutional or individual level. The advantage of developing an improved citation capture tool is that it is likely to respond more adequately to changes in how UDC collections are used, both now

and in the future. An improved citation capture tool is also likely to overcome a number of the other challenges identified above, including a lack of training in how to cite accurately among some users; researchers' ingrained habits; remaining responsive to repository name changes and mergers; and developing a system applicable to both traditional academic publications, online publications and social media platforms.

5.2 Data Sources

There are a number of existing tools designed to capture citations, the two most widely used of which are Data for Research and Google Scholar. Both tools offer an extensive but not yet comprehensive database of published sources, meaning that a combination of both databases is a minimum requirement for capturing citations at the moment. The process currently involves the significant burden of cleaning and analysing citation capture data to ensure that each individual source and/or citation to UDC repositories is counted only once (and not once for each citation capture tool used).

As well as existing citation capture tools, there may also be opportunities to explore partnership working with reference management software companies, such as EndNote and the student plagiarism software company, Turnitin. Both software companies hold extensive databases of published sources that include UDC citations; access to these datasets would facilitate analysis of current UDC citations with minimal burden to inform the development of new software. As noted above, there may also be scope to work closely with ORCID in order to develop unique digital identifiers for UDC institutions based around ORCID's existing model of unique digital identifiers for individual researchers.

There are a number of significant challenges attached to reliance on existing tools and software to capture citations to UDC repositories. These include:

-  **Acknowledging data protection limitations:** Partnership working with both Turnitin and Endnote, as well as other software companies in the sector, may be impeded by data protection legislation, especially with respect to user data. Partnership working may be possible with respect to the broader database of published sources as these are likely to be drawn from existing data in the public domain.
 -  **Supporting UDC repositories:** At present, relatively few repositories actively capture citation data. While there is a general lack of awareness that citation capture is possible for UDC sources, the complexity of capturing citations using existing tools is likely to be a barrier to UDC representatives, many of whom may not have the technical knowledge to capture and analyse the data. User-friendly software is essential to overcoming these challenges.
-

5.3 Bespoke Tool

Even with a more sophisticated approach to using existing data sources, there remains a clear gap in terms of an accessible, user-friendly software platform that is able to extract and analyse citation data with minimal technical expertise required by the software user, e.g. UDC repository staff. As such, a strong case can be made for assessing the feasibility of developing bespoke software that provides a centralised, user-friendly platform to identify and analyse UDC citations. Workshop participants emphasised that processes involved in using a bespoke tool should be as automated as possible; for example, repository staff should be able to press a button and create a graph showing how often their repositories are being cited.

Following this pathway would require significant investment, both financially and in terms of developing the programming needed for the software to run effectively. However, it would also offer the opportunity to create a programme that mitigates against erroneous citations to UDC repositories by academic and other researchers unaware of the appropriate format. This may include allowing for common misspellings of repository names and duplication of three-letter codes by developing the tools needed to distinguish

between different repositories, ideally based on the citation of the full repository name within the initial citation for each publication. Project partners may also be able to explore the possibility of extending the software to include social media and other non-academic publications.

It is important to note that developing a bespoke tool should not be seen as a substitute for engaging with repositories, academics and other stakeholders to improve the accuracy and consistency of citations in the first instance. However, even with improved accuracy in UDC citations, the use of existing data sources does not yet represent a straightforward process for UDC representatives to collate accurate data on citations without a significant investment of time and resources, as well as a working knowledge of collecting and analysing data. As such, there is a clear need to streamline the data collection and analysis side of the citation capture process in order to maximise potential benefits to the UDC sector. There may also be scope to work collaboratively with existing software developers working in related fields, e.g. CrossRef.

6. Implementation Partners

Project partners would need to work closely with other organisations and sectors in order to maximise potential for uptake of the proposed standardised model. Key implementation partners identified through the current consultation include:

- Learned societies and research centres across key disciplinary areas, for example, the Royal Historical Society and Institute for Historical Research.
- Software developers, for example, EndNote, Turnitin and CrossRef.
- Publishers, for example, book and journal editors.

There is likely to be a significant campaign required to engage publishers. Publishers and editors typically work with a high degree of autonomy, with citation guidelines typically informed by the preferred house style. The lack of engagement by publishers and funders across the interviews and surveys may also be indicative of a lack of interest within the sector for a standardised model for UDC citations. As such, there would need to be a concerted effort to demonstrate the benefits of the proposed model for publishers and editors, for example, consistent use of repository citations will save time (and therefore money) in the editing and proofreading process and publishers will also be subject to fewer queries from repositories.

Industry conferences and working groups are two platforms to explore further the potential engagement with key stakeholders within the sector; however, this approach is most likely to be effective once the citation capture project has already achieved significant momentum on its own merits.

Funders, such as the UK Research Councils, were originally identified as potential implementation partners during the scoping phase of the project. The broad consensus from the primary research phase is that specific requirements for citations lie outside their remit. As with publishers, there was also limited uptake from funders during the main research phase of the project.

That said, feedback received from one funder indicates that there may be scope for funders to provide broader support and endorsement of a proposed model, for example, signposting a centralised platform where researchers could find out more about the model, rather than taking an active role in enforcing any proposed model. Workshop participants also noted that some funders, such as The Wellcome Trust, take an active role in embedding best practice through their funding requirement, for example, around open access publications. As such, they were identified as a potential test case for rolling out the citation model in partnership within the funding sector.

Interviewees and workshop participants also noted the benefits of engaging with an international audience, initially to publicise a UK-based model across global users of UK archives but with the potential to roll out

the model internationally in the future. Potential partners include the International Archives Council, Archives Portal Europe, the International Alliance of Research Library Associations (IARLA), and NISO.⁹

7. Conclusions

7.1 Key Findings

- 🌀 Key criteria for the development of a standardised citation model for UDC repositories include integration with existing conventions; accessibility for academics and non-academics; having a single point of access for guidance; and relative brevity of references.
 - 🌀 Possible citation models include reference to the full repository name for the initial citation, followed by an abbreviated letter code based on the repository name for subsequent citations; or a model based on existing ARCHON codes.
 - 🌀 Evidence from primary research points towards three-letter codes as the most appropriate citation model. The principal benefit is that it can be integrated with current practice. While this model would lead to duplication of certain codes, this is not viewed to be a barrier to citation capture across academic publications as long as the full repository name is cited in the initial citation.
 - 🌀 Validation workshop findings indicate that a model based on existing ARCHON codes may offer greater integration with emerging digital infrastructures, such as ORCID codes, as well as ensuring that each repository would have a unique identifier without duplication. Challenges associated with this model include that not all repositories currently have ARCHON codes, and that the model would create a greater emphasis on users knowing and accurately reproducing repository codes.
 - 🌀 Further challenges associated with both models include the ability to adapt to repository name changes; capturing online and social media citations; and the time and financial investment which is likely to be required to ensure that the guidelines are adopted by users.
 - 🌀 Strategies to embed best practice and ensure buy-in for a standardised model include the widespread publication of guidelines; support from publishers; encouraging the uptake of catalogue plug-ins to increase the accuracy and consistency of UDC citations; and the delivery of a publicity campaign to raise awareness and increasing the knowledge and skills required to capture citations.
 - 🌀 Investing in bespoke citation capture tools to automate the citation capture process is likely to be beneficial to ensure accurate and comprehensive data collections; another advantage would be the ability to respond to changes in how UDC collections are used, both now and in the future.
 - 🌀 Challenges associated with using existing data capture tools include data protection limitations which may limit future partnership working; low usage of existing tools among academics (raising questions about the effectiveness of such tools from the user perspective); and the lack of technical knowledge among UDC repositories to capture and analyse the data.
 - 🌀 Key implementation partners identified include learned societies, software companies, publishers and funders. Any partnership working with publishers and editors is likely to need a concerted effort to demonstrate the benefits of a standardised model to their sectors.
-

7.2 Recommendations

The following recommendations reflect the key areas for further development needed in order to achieve the long term goals of the citation capture project: (i) improve accuracy and consistency of UDC citations,

⁹ Potential implementation partners were identified by both research participants and project partners.

and (ii) collate more reliable and effective data on citations by UDC repositories. These recommendations have been developed further in the implementation plan.

Recommendation One: Citation Model

The project partners will need to consider the relative benefits and challenges associated with the two potential citation models identified through this research project: namely, a three-letter code based on repository initials; and unique identifiers based on existing ARCHON codes.

The weighting of these benefits and challenges is likely to centre on the value attached to the key criteria that frame the development of a citation model. While integration with existing conventions and ease of use were emphasised across the primary research, important considerations were raised during the validation workshop around ensuring that any citation model meets the needs of an increasingly technologically proficient workforce across both academic and archival sectors, as well as ease of use in relation to the growing number of both traditional archival resources available online and fully online repositories.

Recommendation Two: Centralised Guidance

The proposed citation model should be accompanied by centralised guidance, for example, through an online user-friendly webpage created and maintained by the project partners. This centralised guidance would be designed to provide repository staff and academics with information on how to implement the citation model; a database of codes that academics and other researchers can easily consult when creating citations to UDC repositories; and an overview of the broader citation capture project including its purpose and benefits for repositories, academics and other key stakeholders.

Recommendation Three: Assess Repository Capability

It is recommended that the project partners engage in a broader consultation process with UDC repositories in order to assess the extent to which staff possess the knowledge and skills required to improve citation capture practices. These include, for example, possessing the subject knowledge and technical expertise to upload citation guidelines to their websites, and using existing data sources to capture and analyse citations to UDC materials held by their repository.

This consultation process would focus on identifying clear strategies to successfully embed best practice within these institutions. One strategy to explore during this consultation process would be the setting up of a train-the-trainer model to support UDC representatives to develop the necessary skills and expertise to ensure best practice is embedded. Such training may include support to develop basic technical skills to run and maintain online best practice guidelines, catalogue plug-ins and citation capture practices.

Recommendation Four: Increasing Knowledge and Awareness

The delivery of a widespread publicity campaign is recommended to increase awareness of the proposed citation model across all potential stakeholder groups, including academics and other researchers, publishers, learned societies and other disciplinary or professional associations. This campaign should focus on informing repositories and their users about the proposed citation model and the project's broader goals. Increasing knowledge and awareness is likely to go hand-in-hand with developing centralised guidance (Recommendation Two) so that individuals have ongoing access to relevant information about the citation model and broader citation capture project.

Delivery of this recommendation would also benefit from more widespread publication of best practice guidelines by UDC repositories, as this in turn may raise awareness of the proposed citation model, as well as encouraging more users — especially academics — to identify repositories as a key source of information on how to cite UDC materials, alongside existing disciplinary norms and publisher guidelines.

Recommendation Five: Improving Software Solutions

We recommend working closely with existing software companies in order to develop a streamlined approach to the process of capturing and analysing citation data. This may include, but not be limited to, initial discussions with ORCID with respect to creating unique digital identifiers for institutions based on their existing unique digital identifiers for individual researchers, as well as approaching software companies such as TurnItIn that hold a comprehensive dataset to explore the potential for using that dataset to capture citation data.

In the longer term, we also recommend that the project partners explore the possibility of developing bespoke citation capture software in order to achieve the long-term goals of the project. As well as overcoming challenges that users have experienced when using existing citation capture tools, this would also create a valuable resource for UDC repositories as well as academics and other researchers.

If delivered alongside a significant publicity campaign to support the implementation of a standardised model (Recommendation Four), the development of improved citation capture tools is likely to make the benefits of following the proposed standardised model more tangible for repositories and their users.

Recommendation Six: Partnership Working

It is recommended that a consultation process is undertaken to assess levels of engagement among learned societies in terms of both the design and publication of the proposed model.

There is a strong likelihood that the project partners could successfully engage with learned societies in key disciplinary areas for UDC materials, given that many of them have a high number of academics in senior positions and that this stakeholder group has been highly engaged in the current phase of consultation. Project partners may wish to engage in a process of stakeholder mapping to identify the most relevant learned societies and, subsequently, to undertake a consultation process with these societies to assess levels of engagement in terms of both the design and publication of the proposed model.

Given the lack of engagement from publishers thus far, project partners may want to consider exploring (i) potential strategies to engage with publishers, and (ii) the feasibility of implementing the standardised model without widespread support from the publishing sector if this engagement process remains unsuccessful. The prospect of limited engagement from publishers further underscores the importance of developing a standardised model for citations that reflects current practices, as this increases the likelihood that academics and other researchers who implement the new model would be able to do so within existing requirements specified by publishers and editors.

8. Appendices

8.1 Desk Review

8.1.1 Current Landscape

Unique and Distinct Collection (UDC) Citations

Unique and distinct collections (UDCs) comprise original, primary sources across a range of medias from cultural and literary heritage to audiovisual materials and born digital content. UDC collections are housed or maintained online by a broad range of institutions, including research libraries, archives, cultural and heritage institutions, and online and digital institutions.¹⁰ Previous research has highlighted the need for UDCs to improve the strategies through which they demonstrate impact in a difficult funding climate, as well as working collaboratively across the sector to increase advocacy¹¹ and enhance collection development.¹²

Developing the use of metrics across the sector is key to achieving these long term strategic goals. Metrics that help to establish the use and value of UDCs focus primarily on data that measures the use of UDC materials, including citations to UDC materials within published research. Early explorations of the ease and effectiveness of tracking citations to UDC materials have found significant variation in the way that research refers to holding institutions, which in turn creates a challenge in terms of how to gather reliable data consistently and effectively through citation capture.¹³ Established research for citations in general indicates that up to 20% of citations in academic publications may be subject to error, although many of these are relatively minor.¹⁴

Further challenges that currently limit the scope for establishing a national citation capture project include the absence of a standardised form of citation to UDC repositories; the lack of consistent advice and guidance provided by UDC repositories; and current citation capture tools do not provide comprehensive records of citations meaning that data needs to be synthesised across multiple data collection sources to capture citations, including but not limited to JSTOR Data for Research and Google Scholar.¹⁵ A further challenge is inherent in increasing digitisation of archival material meaning that any standardised model needs to be applicable to online archive collections.

The key foci of the current research are (a) to explore the scale of the challenge with respect to establishing a standardised form of citation to UDC repositories and (b) to identify key factors for successfully implementing a standardised citation model for UDC repositories. This standardised citation

Unique and Distinct Collections (UDCs)

'A collection that, regardless of format or location within an institution, derives significance from its interest to research, teaching or society through its association with a person, place or topic, such as to distinguish the constituent items from similar items which may exist elsewhere.'

– RLUK UDC Report (2014)

¹⁰ Cullingford, A. et al. (2014). *Unique and Distinct Collections: Opportunities for Research Libraries*. London: RLUK.

¹¹ Cullingford, A. et al. (2014). *Unique and Distinct Collections: Opportunities for Research Libraries*. London: RLUK.

¹² Dooley, J. et al. (2013). *Survey of Special Collections and Archives in the United Kingdom and Ireland*. Dublin, Ohio: OCLC Research.

¹³ Internal Programme Documentation - Citation Capture Review II

¹⁴ <https://scholarlykitchen.sspnet.org/2014/11/06/why-are-publishers-and-editors-wasting-time-formatting-citations/>

¹⁵ Internal Programme Documentation - Citation Capture Review II

model would be user friendly and easily integrated into current citation practices; the working assumption is that this model would complement the current system of numeric ARCHON codes used at a systems level to catalogue and identify archival collections across the UK and relevant international repositories.¹⁶

While there is no current sector-wide guidance for UDC repositories that provides best practice for citation capture, best practice recently developed for the American archive sector identified a number of key guidelines to improve citation capture. These include establishing which types of publications to monitor citations across both academic and non-academic platforms; requiring users to report cited material in their publications through user registration terms, call slips, permission-to-publish agreements, and/or reproduction order forms; and developing strategies to keep an up-to-date record of citations through key word searches and the use of citation capture tools, for example, by setting up Google search alerts for repository name or specific collection names.¹⁷

It should also be noted that there are complementary projects designed to standardise citations to specific types of material from rare books¹⁸ to digital software.¹⁹ While these standardisation projects focus on how to cite specific items rather than the institution responding for preserving or maintaining that item, these projects indicate cross-sector engagement to improve best practice.

Digital Tools and Software

Two leading sources for citation capture are JSTOR Data for Research (dfr.jstor.org) and Google Scholar. The former provides data on articles held within the JSTOR repositories, while the latter has a much wider remit but is limited to sources that have been digitised in some format.²⁰ Data for Research is currently in BETA format. It provides free access to bibliographic data, although a subscription is needed to view the full source.²¹ Other sources of citation capture for academic publications include Elsevier's Scopus. Turnitin may also provide insights into student usage.²²

There may also be potential to make use of data held by reference software companies in order to facilitate citation capture. The two key software companies in this respect would be the reference management software EndNote (and similar competitors) and the plagiarism detection service Turnitin. EndNote allows users build their own library of citations that can then be imported into documents via a plug-in.²³ The potential exists to draw relevant citation data based on citations to UDC materials imported into users' libraries and/or exported into documents; however, there is limited information in the public domain to assess whether EndNote would be receptive to collaborative working, or whether they collect and hold this data.²⁴

Turnitin functions by comparing submitted work to its database of over 1 billion student papers, as well as 65 billion indexed webpages and 170 million scholarly publications and articles. Turnitin's vast database makes it a good potential source of data for citation usage; however, access would likely to be restricted to the database of sources along as the company stated that sale of users' data to third parties would violate

¹⁶ <http://www.nationalarchives.gov.uk/help-with-your-research/discovery-help/find-an-archive>

¹⁷ Internal Programme Documentation - SAA-ACRL/RBMS, Standardised Statistical Measures and Metrics

¹⁸ <https://rbms.info/scf/introduction/>

¹⁹ <https://research-software.org/citation/>

²⁰ Internal Programme Documentation - Citation Capture Review II

²¹ <http://www.jstor.org/dfr/>

²² Internal Programme Documentation - Citation Capture Review II

²³ See the list of currently compatible sources here: http://www.myendnoteweb.com/help/en_us/ENW/he_capture.htm

²⁴ It is not included in EndNotes standard license agreement. See <https://endnote.com/license>

its own privacy policy and interpretation of GDPR requirements.²⁵

Further tools that may be used by both individual researchers and research organisations to track citations includes ORCID persistent identifiers²⁶ and International Standard Name Identifier (ISO 27729).²⁷ Both these models and other similar initiatives²⁸ are designed to improve data capture and analysis. There is also scope to explore existing tools and standards related to the metadata for archival and UDC materials held within or uploaded to digital archives; for example, the Metadata Encoding and Transmission Standard (METS) is designed to improve digital collection development for digital library holdings through standardised metadata schema.²⁹ Both this and other similar projects may have implications for collecting and analysing usage and citations related to digitised materials.

Other models and software that may provide avenues for further research to improve citation capture tools include a parsing model to extract relevant bibliographic data that allows for variations in content and style,³⁰ as well as software to draw additional information from citations including identifying secondary citations.³¹ There are also increasing software options to assist in the digitisation and presentation of archival materials led by the Open Preservation Foundation.³²

8.1.2 Mapping Citation Guidance

Introduction

Most of the guides and commentaries on citation do not mention the institution holding a cited item in a UDC. Institutions offering guidance on citation refer students and researchers to one of a handful of guides (e.g. Harvard, Chicago, University of Western Australia and APA) but with little if any, mention of the name of the institution; instead the focus is on the type of material to be cited. There is also significant variation in the proposed structure and style of citations, as shown in the examples below.

Citation Guidance Structure

A significant number of UDC repositories provide online guidance on how to cite materials held in their collections. A light-touch survey of online citation guidance provided by UDC repositories indicates that there is a broad consensus on the key components required for a full citation:

- **Name of Repository:** This typically includes any preferred abbreviated format for referring to the repository or institution, e.g. The National Archives (TNA).
- **Catalogue Reference:** This typically comprises an alphanumeric code used by the archive to refer to each item held within their collections. This may be made up of a number of different elements, and varies

²⁵ https://guides.turnitin.com/Privacy_and_Security It is also important to note that Turnitin is subject to multiple controversies within academic communities, principally on issues around privacy and copyright-violation, which may bear upon its suitability as a partner organisation.

²⁶ <https://orcid.org>

²⁷ <http://www.isni.org>

²⁸ Fenner, M. et. al. (2015). *D2.1: Artefact, Contributor, and Organisation Relationship Data Scheme*. *THOR: Technical and Human Infrastructure for Open Research*.

²⁹ <http://www.loc.gov/standards/mets/>

³⁰ Sautter, G. And Böhm, K. (2014). *Improved Bibliographic Reference Parsing Based On Repeated Patterns*. *International Journal on Digital Libraries* 14:1-2.

³¹ <http://blogs.plos.org/tech/rich-citations/>

³² <http://openpreservation.org>

considerably across different institutions. For example, the National Archives requires users to provide details of the department code, series number, piece number and item number (e.g. E 134/25Eliz/Trin I).

- **Internal Identifier:** This refers to the folio, page, docket, membrane or other number specific to the item itself, e.g. referring to pages using p. or pp., f. or ff. for folios and m. or mm. for membranes.

It is also common to see the following reference components recommended to provide clarity:

- **Archive Collection:** This refers to details of the specific archive collection. Specifying the archive collection is important for copyright purposes when collections have been donated by an external source. From a user's perspective, it also helps readers to identify the specific collection held by a large UDC repository, e.g. 'J. Lyons and Company Collection Ltd' (an archive collection held by London Metropolitan Archives).
- **Description of Item:** This is often recommended to help readers understand the type of item referred to, e.g. 'Letter from John Smith to Mary Smith', or 'Journal of the House of Commons'.

As noted in the analysis section above, the level of detail and ordering of these components varies across different UDC repositories. The example citations detailed below provide an indication of the format recommended by leading UK and international archives.

Example Citations from UK Repositories

The following colour code is used to refer to each of the components listed above

Name of Repository
Catalogue Reference
Internal Identifier
Archive Collection
Description of Item

National Archives³³

The National Archives (TNA): C 139 Chancery: Inquisitions Post Mortem, Series 1, Henry VI
The National Archives of the UK (TNA): ADM 1/2233/19

For both of these example citations, the repository name can be abbreviated to TNA after giving the full repository name in the first reference. The National Archives does not provide reference information to specific items within their standard catalogue reference; however, guidance suggests that users may wish to provide additional descriptive information to identify individual items, as required.

³³ <http://www.nationalarchives.gov.uk/help-with-your-research/citing-documents-national-archives/>

London Metropolitan Archives³⁴

London Metropolitan Archives, City of London LCC/TWYS/GEN/01/00, from the London County Council collection

Woburn: agent's policy register 1843 -1895, London Metropolitan Archives, City of London CLC/B/192/F/017/MS38866, from the Royal and Sun Alliance Insurance Group collection, Sun Insurance Office Limited.

The London Metropolitan Archives also specify that the archive can be referred to in the abbreviated format LMA for both initial and subsequent references.

Parliamentary Archives³⁵

Parliamentary Archives, HL/BR/1/451

Parliamentary Archives' citation guidance refers to archive material held by the institution. Printed acts, printed journals, official reports and other published materials should be cited using standard citations as these items are not unique to the archive.

Modern Records Centre, University of Warwick³⁶

Letter from Harold Macmillan to Richard Crossman, 15 December 1943; Richard Crossman Archive, Modern Records Centre, University of Warwick (MSS.154/3/AU/1/96)

University of Leicester³⁷

University of Leicester: Archives and Special Collections, MS237/1/2/1/1 pp.24-26. 'Scrapbook containing theatre programmes, articles and review of 'Entertaining Mr Sloane', collected and compiled by Joe Orton, 1964-1967'.

The University of Leicester's citation guidance also provides examples from other institutions; this guidance is notable for following their internal citation model rather than the format recommended by the institutions themselves.

The National Archives, FO 1060/262. 'Documents submitted to the Council of Enquiry on Restitution'.

British Library, MS48927 ff.1-54. 'D'Abernon Papers, internal embassy memoranda at Berlin, 1920-5'.

³⁴ <https://www.cityoflondon.gov.uk/things-to-do/london-metropolitan-archives/about/Documents/guideline-citation-of-lma-documents.pdf>

³⁵ <https://www.parliament.uk/business/publications/parliamentary-archives/visiting-and-services/copyright-and-use/>

³⁶ <https://warwick.ac.uk/services/library/mrc/using/citation/>

³⁷ <https://www2.le.ac.uk/library/help/referencing/footnote/footnote#Manuscripts>

Examples from Online Repositories

Old Bailey Proceedings Online³⁸

Old Bailey Proceedings Online (www.oldbaileyonline.org, version 6.0, 17 April 2011), April 1754, trial of Elizabeth Canning (t17540424-60).

The Old Bailey Proceedings Online provide digitised copies of original archival material alongside a full transcription of the contents. Each separate entry has a 'cite this text' widget that pops open to give a full citation for each specific item. There is also separate citation guidance within the website's copyright section that sets out the preferred format for citations. In terms of the repository name, citation guidelines state that the repository name can be abbreviated from Old Bailey Proceedings Online to OBP in subsequent references.³⁹

London Lives⁴⁰

Ordinary of Newgate's Account, London Lives, 1690-1800, OA17260627 (www.londonlives.org, version 1.1, 17 June 2012), 27 June 1726.

St Martin's in the Fields Pauper Biographies Project, Settlement Examinations, Ann Carter, 22 February 1733, London Lives, smdsset_19_1930 (www.londonlives.org, version 1.1, 17 June 2012), Westminster Archives Centre, Ms. F5025.

London Lives is an online collection of digitised archival material from eight different London libraries supplemented by datasets from a further fifteen UK and international sources. Each entry includes a digitised copy of the original archival material alongside a full transcription of the text. Bibliographic details for the originating archive, date and source of the manuscript material, and the London Lives catalogue reference are included for each entry. A separate citation guide provides details of the recommended format for creating citations depending on the originating source; this guidance states that researchers may abbreviate the repository name to LL in subsequent citations.⁴¹

Examples from International Repositories

National Archives of Australia⁴²

National Archives of Australia: Antarctic Division; P1556, Files containing Antarctic Station Reports with Station Log Books interspersed, chronological series, 1947-; Heard Island 1954, Heard Island 1954 Seal Sightings Log, 1954-55.

³⁸ <https://www.oldbaileyonline.org/index.jsp>

³⁹ <https://www.oldbaileyonline.org/static/Legal-info.jsp#citationguide>

⁴⁰ <https://www.londonlives.org>

⁴¹ <https://www.londonlives.org/static/Legal.jsp>

⁴² <http://www.naa.gov.au/collection/fact-sheets/fs07.aspx>

The National Archives of Australia recommends, however, that users employ an abbreviated format for references specifying only the 'series' and 'control' number from the full catalogue reference and recording the repository name in its abbreviated form, as follows:

NAA: P1556, Heard Island 1954.

National Library of Australia⁴³

Minute Books, Records of the Australian National Research Council, National Library of Australia, MS 482A/1/4-5

The National Library of Australia also provides an example of a suitable abbreviated reference to be used following the initial reference to the item. This includes an abbreviated reference to the archive as NLA followed by the catalogue reference, as follows:

NLA MS 482A/1/4-5

Purdue University Libraries⁴⁴

Purdue University Library provides examples of how to cite its materials for different citation models, including the Chicago model as follows:

Booth Tarkington to George Ade, 8 May 1924, Box 10, Folder 5, George Ade Papers 1878-2007, Purdue University Archives and Special Collections, Purdue University Libraries.

Purdue Online Writing Lab⁴⁵ suggests following guidelines provided by repositories, or adopting the standard format recommended by the university library in the absence of such guidelines with added fields for location and date accessed.

National Archives of the United States⁴⁶

Amos T. Ackerman, Attorney General, to Richard Crowley, U.S. Attorney, New York, Nov. 23, 1871; Vol. C, Oct. 27, 1871–Apr. 23, 1873, p. 60; Instruction Books, 1870–1904; General Records, Letters Sent, 1849–1919; General Records of the Department of Justice, Record Group 60; National Archives Building, Washington, DC.

Citation guidance states that the repository name can be abbreviated to NAB for subsequent references, as well as recommended abbreviations for other elements of the citation including the catalogue reference and collection name.

⁴³ <https://www.nla.gov.au/how-do-i-acknowledge-the-library-in-my-publication>

⁴⁴ <http://guides.lib.purdue.edu/c.php?g=352889&p=2378064>

⁴⁵ <https://owl.english.purdue.edu/owl/resource/988/06/>

⁴⁶ <https://www.archives.gov/files/publications/general-info-leaflets/17-citing-records.pdf>

Massachusetts Historical Society⁴⁷

Fitzjohn Smith to Impatience Jones, Nov. 3, 1634, Fitzjohn Smith Papers, Edifice Historical Library, Providence, R.I.

Guidance notes include recommendations to use appropriate abbreviations for both collections and repository, for example, abbreviating Fitzjohn Smith Papers to FSP and Edifice Historical Library to EHL in subsequent citations.

The Massachusetts Historical Society is also notable for discouraging the use of catalogue reference for archival and special collection materials unless specified in repository guidelines on the basis that these catalogue references are liable to change if the collection is reprocessed.

8.1.3 Citation Guidance: Trends Analysis

Citations to Repositories

-  **Use of three-letter repository codes:** A number of repositories recommend using a three-letter code; however, a significant number of repositories do not specify a preferred abbreviation in their citation guidance. Recommendations for the use of abbreviated repository codes also vary, with The National Archives recommending their use in both the initial citation of the repository name (e.g. *The National Archives (TNA):WO 32*) and then without the full repository name for subsequent abbreviated references (*TNA:WO 32*), whereas the London Metropolitan Archives requests that the initial citation should be listed in full without the code (e.g. *London Metropolitan Archives, City of London CLC/B/074/MS00817*) with the abbreviation LMA only used for subsequent citations. Three-letter abbreviations are also used by some of the larger international repositories included in this review, including the National Archives of Australia (NAA) and the National Library of Australia (NLA).
-  **Citation to online repositories:** Citation guidance provided by online archival collections identifies the online repository as the principal repository name with the originating archive included within citations where appropriate. Best practice identified for online repositories includes widgets for each separate entry to increase likelihood of users adopting the preferred citation style. The two examples included in the analysis above also recommend the use of abbreviated repository references for subsequent citations.
-  **Use of ARCHON repository codes:** ARCHON codes has the potential to provide a unique reference code for all repositories in the UK. At present, only publicly funded archives have an ARCHON code except for private archives that have opted in. While The National Archives lists ARCHON codes on its website's 'Find an archive' section,⁴⁸ none of the citation guidance reviewed so far has advised the use of ARCHON codes. This is likely because ARCHON codes are used internally by archives for external data and systems analysis rather than for citations and other user-based activities.
-  **Capturing historic citations:** London Metropolitan Archives specifies that predecessor repositories now held by the LMA, e.g. Guildhall Library Manuscripts, should not be referenced. It is recommended instead that the name of repository is referenced as LMA. While this represents best practice for current citations, it raises the issues of how to capture historic citations to old repositories whose collections have been taken over by a larger repository.

⁴⁷ <http://www.masshist.org/publications/styles/mhr-archCites>

⁴⁸ <http://discovery.nationalarchives.gov.uk/find-an-archive>

General Trends in Citation Guidance

- 🌀 **Extent of guidance:** The level of detail given in citation guides varies greatly across both UK-based and international repositories. For example, both The National Archives of the UK⁴⁹ and the National Archives of Australia⁵⁰ provide examples of how citations should be presented for a number of different types of material, as well as providing details about how each element of the catalogue reference is constructed, whereas The Modern Records Centre at the University of Warwick⁵¹ give only their 'bare minimum' and 'recommended' formats with examples of the latter.
- 🌀 **Structure of recommended citation style:** A key theme noted throughout the guidance reviewed was a lack of consistency in the structure of UDC repository citations. While most archives use the same component parts for citations, the ordering of components and level of detail required varies significantly across repository guidelines. This is clearly illustrated in the example references drawn from citation guidelines above.
- 🌀 **Variation in use of catalogue references:** Each of the repositories included within this review uses its own internal system to catalogue UDC collection materials. As well as variation in the specific codes used, there is also significant variation in the terminology used to refer to catalogue references; for example, The National Archives use department code, series number, piece number and item number to make up their catalogue reference, while other archives refer to the structure of their catalogue references in terms of document reference number, name of series, and so forth.
- 🌀 **University affiliated archives:** Citation guidance provided by research libraries affiliated to universities is more likely to provide examples of how to format archival references according to different citation styles. For example, Purdue University Library provides guidance on how to cite archival material held within their repository formatted for the American Psychological Association (APA), the Modern Language Association (MLA) and the Chicago Manual of Style.⁵² There are also examples of university affiliated archives providing students with additional guidance on how to reference UDC material held by other repositories; for example, additional citation guidance provided by the University of Leicester indicates how to reference material from its own special collections as well as how to apply that same guidance to items held by the National Archives and the British Library.
- 🌀 **Copyright and archive collections:** While relatively few repositories recommend including the name of the specific archive collection within the reference, including this information within the reference may be relevant for copyright. For example, London Metropolitan Archives states that this is to ascertain the copyright owner as many of their collections remain in the ownership of the depositor. Similarly, for the Parliamentary Archives, the guidance notes that copyright for an item may be owned by someone other than the repository but unlike LMA, this ownership information is not integrated into the citation format.
- 🌀 **Referencing digital materials:** Relatively few repositories provided guidance on how to cite digital materials, which may be indicative of wider trends. Of those that did provide this guidance, UCL Library Services states that, wherever possible, a Digital Object Identifier (DOI) should be included in the citation.⁵³ DOIs ensure that data can be linked, even if the location changes (i.e. if the URL of a digital record changes). DOIs are an internationally recognised and supported standard. They can also be applied to physical objects and are not limited to open source data.⁵⁴ In terms of international examples of best practice, Purdue University's Online Writing Lab specifies the use of DOIs or other digital

⁴⁹ <http://www.nationalarchives.gov.uk/help-with-your-research/citing-documents-national-archives/>

⁵⁰ <http://www.naa.gov.au/collection/fact-sheets/fs07.aspx>

⁵¹ <https://warwick.ac.uk/services/library/mrc/using/citation/>

⁵² <http://guides.lib.purdue.edu/c.php?g=352889&p=2378064>

⁵³ <http://www.ucl.ac.uk/library/digital-collections/citing>

⁵⁴ <https://www.bl.uk/datacite/overview/does-and-citation>

identifiers such as Archival Resource Keys (ARKs) within citations. They also provide a structure for the citations to digital materials but state that this should only be used if the holding repository does not have its own citation guidelines.⁵⁵

8.1.4 Social Media Platforms

A selection of social media outlets were sampled to identify current online citation practices. In order to gauge practices across different audience profiles, a range of outlets were selected including the blogs of high-quality academic-style news/blog sites, academic institutes and repositories themselves. Twitter was also selected as a populist indicator of social media trends.

Academic Blogs

None of the blog posts sampled contained full bibliographies at the end of the article; all relied on in-text citation, the format of which varied depending on the publication and author. For example, citations in articles on *The Conversation*^{56/57} were often in the form of hyperlinks to other sources; this includes citations to other articles and publications, archive catalogues, journal abstracts, full journal articles, book synopses (on publisher/retailer websites or Google Books), digitised books and digitised primary sources. The anchor text for these links were in-text rather than in the form of a citation-style superscript. Furthermore, until it was clicked on, it was unclear whether the link was intended as a formal citation or to provide general information. All of the sampled articles from *The Conversation* used citations that were digitally linked to another online source.

In contrast, sampled blog posts from the Institute of Historical Research used traditional abbreviated citations in text. In some posts such as 'From Roman villa to Brutalist architecture – homes, housing and households in the BBH', the citations were used without hyperlinks to digital resources, implying that they were only used when a link to a digital resource was either unavailable or had not been sourced by the article author. However, in other posts, such as 'Home in the IHR library: inventories in the library collection', the abbreviated citations given did link to further resources related to the citation.

One key observation from the sampled blogs was that there is variation in citation style between articles on the same site. This is possibly because the sites have contributions posted by multiple authors who have used different styles which have not been standardised before publication. For example, the two IHR blog posts referenced above were written by different authors, which may account for the stylistic differences in the citation. Similarly, while many of the citations on *The National Archives* blog took the general form of *item description, catalogue number*, e.g. Archaeological find, 1949-1950 (WORK 25/196),⁵⁸ others were listed in a different format, e.g. AR 1/528 11-18. Suffragettes: descriptions and photographs.⁵⁹

Additionally, in one blog post with multiple contributing authors, it was noted that citation styles varied within the article. In 'Suffragettes, 1912: 'Rather broken windows than broken promises'',⁶⁰ citations varied between *catalogue number, item description*, e.g. DPPI-23 Ex142 Letter from Amy Woodburn to Mrs Pankhurst 26 February 1912, and *item description, catalogue number*, e.g. Summary of claims, damage by Suffragettes, March 1912. Catalogue reference MEPO3-1787.

⁵⁵ <https://owl.english.purdue.edu/owl/resource/988/07/>

⁵⁶ <https://theconversation.com/a-brief-history-of-dyslexia-and-the-role-women-played-in-getting-it-recognised-89055>

⁵⁷ <https://theconversation.com/drilling-holes-in-the-skull-was-never-a-migraine-cure-heres-why-it-was-long-thought-to-be-90782>

⁵⁸ <http://blog.nationalarchives.gov.uk/blog/stories-tell-archives-civilisation/>

⁵⁹ <http://blog.nationalarchives.gov.uk/blog/international-womens-day-2018/>

⁶⁰ <http://blog.nationalarchives.gov.uk/blog/rather-broken-windows-broken-promises/>

Twitter

Many of the Twitter accounts surveyed did not provide direct citations for images and other document types included in their tweets. For example, the Twitter feed for History Today⁶¹ and The Conversation's UK site⁶² link directly to articles on their main site.



Similarly, The National Archives generally provides links to blog posts or news items on its website rather than providing a formal citation for the item featured in the tweet, although there are instances on their feed of links directly to its image library.⁶³ It should be noted that the image included in the tweet was not always found in the article to which the tweet linked,⁶⁴ so it would very difficult to capture the citation of the image.



Some of the sampled Twitter accounts had tweets which contained details of digitised documents; many of

⁶¹ <https://twitter.com/HistoryToday>

⁶² <https://twitter.com/ConversationUK>

⁶³ <https://twitter.com/UkNatArchives/status/971460838665187329>

⁶⁴ <https://twitter.com/UkNatArchives/status/97177898293153792>

these tweets contained a hyperlink to the digital item itself. This practice was particularly prevalent on accounts belonging to digital collection specialists such as Bodleian Digital Library⁶⁵ and Cambridge Digital Library,⁶⁶ but was also noted on accounts of specific projects run by institutions such as the British Library, for example the British Library Hebrew Project.⁶⁷

When linking back to other websites, some of the links given use link-shortening services such as [bit.ly](#). This does not seem to be done consistently, however, with tweets on The Conversation UK's feed switching between listing full urls, using Twitter's auto-generated [t.co](#) shortening service and using a bespoke shortening service such as [bit.ly](#). The use of a bespoke shortening service could potentially aid with citation capture as it allows for the monitoring of link clicks and other analytics. They could also aid in the tracking and capture of citations contained in retweets of the original material.

8.1.5 Implications for Citation Capture

The following implications have been identified for the Citation Capture project based on the evidence analysed for the literature review:

- 🌀 **Digital tools and software:** The most commonly used citation capture tools are JSTOR Data for Research and Google Scholar; other citation capture tools include Thomsons Reuter's ISI, Elsevier's Scopus and Turnitin. None of these tools provide comprehensive citation data, meaning that analysts are required to synthesis citation data captured across these tools. Further digital tools and software may provide significant opportunities to improve metadata and persistent identifiers for digitised materials and/or traditional UDC materials included within online catalogues; however, further research conducted by technical specialists would be needed to fully scope these opportunities.
- 🌀 **Current citation guidance:** Analysis of current citation guidance indicates a wide variation in the format and level of detail required by UDC repositories; however, almost all citation guidance requires users to include both the repository name and relevant catalogue reference as core components of any accurate citation. As such, all accurate citations should include the repository name in either a full or abbreviated format which supports the underlying hypothesis of the project that that the repository name is the most appropriate unit of data to capture citation data.
- 🌀 **Standardising citations to repository names:** Repositories typically request that the repository name is cited in full within an initial reference. A small number of high profile repositories suggest a preferred three-letter code to be used for subsequent citations, typically a three letter code based on the full repository name, for example, The National Archives is abbreviated to TNA. This may indicate that support for a standardised model based on three-letter codes is likely to integrate relatively easily with existing practices across the UDC sector. There is no evidence of UDC repositories recommending the use of ARCHON codes within citations.
- 🌀 **Adapting to the digital era:** There is a growing trend for traditional paper-based UDCs to be digitised and presented within online repositories. The development of a standardised model for repositories names would need to be suitable for these online repositories as well as traditional UDC repositories.
- 🌀 **Capturing citations across social media platforms:** Citations included within articles and other outputs on social media platforms typically appear to be hyperlinks embedded within the text rather than full citations; there were a number of exceptions to this general trend on scholarly platforms such as the IHR blog. On more popular platforms, such as Twitter, citations were often significantly abbreviated with both [bit.ly](#) and [t.co](#) link-shortening services commonly used. This may suggest that

⁶⁵ <https://twitter.com/BDLSS>

⁶⁶ <https://twitter.com/CamDigLib>

⁶⁷ https://twitter.com/BL_HebrewMSS/status/972797030165860352

capturing citations on social media platforms will be dependent on improving citation capture tools to analyse domain names rather than relying on users to include a standardised reference to the repository name within their posts.

8.2 Interview Analysis

8.2.1 Citation Model Criteria

The key criteria for assessing citation models for academics is that the model is intuitive, uses existing conventions, and provides sufficient detail to allow readers to follow up and consult the source for themselves. Brevity was also a concern for academics working within specific word counts and/or to other requirements set by publishers. Abbreviated references were common following the first instance that UDC materials are cited; as one interviewed explained: *'It becomes repetitive to use the whole name in each citation and, from a publishers point of view, it looks quite unsightly.'*⁶⁸

Academics also highlighted the need for a model that is sufficiently flexible to allow them to adapt specific details to meet publisher requirements; and that avoids them having to look up additional information (e.g. standardised abbreviations) in order to create citations. With respect to social media platforms, the brevity and durability of the model were identified as key criteria for successful uptake with most academics keen to see readable, persistent URLs or other digital identifiers used for citations.⁶⁹

The main criteria among UDC representatives for assessing different citation styles is the ability to locate the source material: *'We do have examples of people coming into the service with a book that just says the document is at our repository. Where we go from there is very difficult.'* Other common responses include accurate citation of the repository name (whether using the full name, abbreviations or repository codes), consistency of citation practices, and the extent to which the citation style is easy to understand and use.⁷⁰

For project partners, length was seen as the most important criteria for assessing the viability of a standardised model; according to one interviewee, short abbreviations are increasingly the norm within academic citations. Both interviewees agreed that it would be significantly easier to develop a standardised model for repository names than for other elements of UDC citations, such as the format of catalogue references and item codes.⁷¹

8.2.2 Current Practices and Conventions

UDC repositories all had preferred ways of being cited, typically using the full repository name. A small number of interviewees referred to common abbreviations used to cite their repository; all of these abbreviated references comprised a three- or four-letter code based on the initials of the repository name. A minority of repositories prefer that the full repository name is used in all citations. Interviewees were also aware that their repository's preferred abbreviation may be shared by other repositories across the UK. As one interviewee explained, *'We use the abbreviation DRO but of course there's lots of other DROs nationwide for the 'D' counties so depends on context whether it's appropriate.'*⁷²

The two principal factors that determine how academics structure citations are disciplinary norms and citation guidance from journal and book publishers. Most academics develop their understanding of disciplinary norms through reading publications by others in their field. As one interview highlighted, *'I want what I produce to be useful to people in my discipline so if there is a set of conventions then I try to follow them.'* Other interviewees noted that not all of their colleagues were sufficiently precise in citations to UDC materials, which limits the ability of other academics to follow up and consult materials cited within their

⁶⁸ Interviews with academics.

⁶⁹ Interviews with academics.

⁷⁰ Interviews with UDC representatives.

⁷¹ Interviews with project partners.

⁷² Interviews with UDC representatives.

colleagues' publications.⁷³

Academics typically introduce abbreviated references to UDC repositories after the first instance of citing that repository; these were usually specific to each individual publication. Three letter codes appear to be the most commonly used approach by academics when abbreviating UDC repository names. While most abbreviated references are developed ad hoc and/or informed by disciplinary norms, a small number of interviewees referred to online citation guidelines produced by UDC repositories or contacted repositories directly to confirm preferred citation style.⁷⁴

As one interviewee highlights, however, there can often be significant variation in the abbreviations used by academics: *'For instance, when referring to the Churchill Archive Centre in Cambridge, I abbreviated it to CAC... [but] I've seen it referred to in many different forms by other academic authors.'* A number of academics emphasised the importance of clearly explaining the use of any abbreviated references within citations in order to circumvent this issue, for example, including a list of abbreviations within the preface or bibliography of a published book.⁷⁵

A high number of repositories had experienced one or more name changes within recent years following mergers or restructuring; as a result, many of these repositories were aware that users cited material using previous repositories names even if a significant publicity campaign had taken place. Repositories that had experienced name changes or mergers in recent years were also more likely to highlight the importance of a flexible system for citations as repositories often adapt letter codes to reflect their current name; these repositories also noted that a comprehensive citation capture system would need to be able to capture both current and historic citations.⁷⁶

A small number of UDC representatives and one project partner felt that ARCHON codes could be used to create a standardised model since it offered an existing system of unique identifiers that covers all UDC repositories.⁷⁷ UDC representatives noted, however, that this approach would have its own challenges as (i) there is a lack of familiarity amongst both researchers and their readership (ii) the burden would be on researchers and their readership to look up codes because the codes are not easily recognisable and (iii) having a clearly recognisable name in the citation helps to popularise individual repositories. It is also notable that no academics taking part in interviews referred to using ARCHON codes nor indicated that these codes would be the suitable basis for a standardised citation model.⁷⁸

None of the interviewees were aware of an existing international citation model relating to UDC repositories; one interviewee responsible for delivering a module on archival method and metadata noted that she had enhanced her understanding of different global approaches to citations by working closely with international students, many of whom are informed by national or regional citation practices.⁷⁹

8.2.3 Social Media Citations

Evidence from interviews with academics indicates that there may be significant barriers to standardising citations on social media platforms, not least because academics rarely use full citations on these platforms. Information included within social media citations of UDC materials is typically limited to the name of the UDC repository and/or a weblink to the relevant materials where appropriate. As one interviewee explained, *'if the point of a citation is to allow the reader to follow it up themselves, it wouldn't be expected for that*

⁷³ Interviews with academics.

⁷⁴ Interviews with academics.

⁷⁵ Interviews with academics.

⁷⁶ Interviews with UDC representatives.

⁷⁷ Interviews with UDC representatives and project partners.

⁷⁸ Interviews with UDC representatives.

⁷⁹ Interviews with academics.

to happen for a public audience.’ Other interviewees noted that they had been actively discouraged from using full citations in order to ensure that online content remained accessible to a broad audience. There were also examples of academics using the web archive version of URLs or DOIs to ensure that weblinks would not become outdated but this was not considered to be standard practice.⁸⁰

UDC representatives also felt that social media outputs operated according to different rules compared with traditional academic publications: ‘Social media is more informal and fast paced. People are less likely to follow rules when quickly publishing a tweet’. One interviewee observed that repositories were limited in how much they could influence users to reference their name on social media: ‘encouragement rather than expectation as [the latter] goes against the current ethos of social media which is in the moment’. Potential solutions identified by the interviewees to capture social media citations include using widgets to develop a short form URL which links to a formalised citation elsewhere; watermarking digital records; or using hashtags to reference material from institutions.⁸¹

Project partners also agreed that a permanent URL is the most appropriate way to cite UDC materials on social media platforms, for example, sharing a weblink to the relevant catalogue entry. According to one interviewee, the challenge is to ensure that these URLs are genuinely permanent so that the references continue to be valuable to scholars working in decades to come. Another interviewee noted that using URLs attached to the repository’s website can be easily tracked from a citation capture point of view because the domain name will be uniquely linked to that repository. Project partners also referred to a particular challenge that may arise with respect to social media outputs, given that academics often put up ‘teasers’ of their research materials, and as such, may be reluctant to provide a full citation before writing up the material for an academic publication.⁸²

8.2.4 Embedding Best Practice

Evidence across the interviews indicates that there is also scope to improve current citation practices through raising awareness around best practice for citations in general, as well as promoting the proposed standardised model.

Over two thirds of UDC representatives state that their repository provides guidelines on how to cite material; however, a high proportion of interviewees explained that this guidance is provided on request rather than being widely available, while other interviewees noted that the guidance on their website was extremely brief and/or difficult to find. Of those that did provide comprehensive guidelines, one interviewee identified Archive Hub as an effective tool to support users to create accurate citations.⁸³

There is also some evidence to suggest that not all academics are aware of existing citation guidance; for example, one interviewee referred to the lack of guidance at The National Archives. This may suggest that UDC repositories would need to put significant resources into ensuring that users are more widely aware of their citation guidelines.⁸⁴

There is clear need for guidelines providing the key information to produce accurate citations to become more accessible. Interviewees across stakeholder groups observed that many academics do not produce accurate citations to UDC materials and that these inaccuracies may be replicated within a new standardised system.⁸⁵ As one project partner interviewee noted, ‘People who can’t cite in full are unlikely to be

⁸⁰ Interviews with academics.

⁸¹ Interviews with UDC representatives.

⁸² Interviews with project partners.

⁸³ Interviews with UDC representatives.

⁸⁴ Interviews with academics.

⁸⁵ Interviews with project partners, academics and UDC representatives.

able to cite to produce a repository code.⁸⁶

UDC representatives also noted that many of their users do not come from academic backgrounds; as such, they often do not have the necessary training to know how to cite accurately and/or may be less likely to be aware of existing citation guidelines and models: *'it would be people understanding how to do it and how to make citations who may not have a background in academic work.'*⁸⁷

While it was widely noted that UDC repositories have an important role to play in raising awareness of any proposed model citation guidance, as well as best practice in UDC citations more broadly, one project partner observed that many repositories have limited technical ability and/or resources to upload an online style guide their website, for example.⁸⁸

Alongside improved guidelines, interviewees across all stakeholder groups emphasised the need for a comprehensive publicity campaign in order for any proposed model citation guidance to become embedded in practice. Interviewees warned against taking a *'heavy handed'* approach to enforcing implementation, preferring instead to see the benefits of the proposed model to be communicated in order to encourage uptake: one interviewee suggested that this could take the form of *'do this and help your archive'* in order to offset the additional effort required to (potentially) look up and use new repository codes.⁸⁹ One publisher also highlighted the need for any publicity campaign to remain an ongoing process: *'just because you have a certain proportion of the market you just want to stick with the message.'*⁹⁰

A number of interviewees also emphasised the importance of linking up a publicity campaign with broader stakeholder engagement, for example, through consultation workshops or appointing regional champions. This was felt to be important so that any proposed citation model guidance would not be perceived as imposed from outside.⁹¹

8.2.5 Citation Capture

Only one of the UDC representatives stated that they systematically capture citation data; this is largely due to the interviewee's awareness that it is possible to look for citations using JSTOR and Google Scholar. Another interviewee stated that capturing citation data had been trialled briefly and then halted because of difficulties in identifying citations even if using a number of different search terms.⁹²

Other UDC representatives mentioned that they become aware of citations on an ad hoc basis, for example, if users contact them for publication rights or to share a copy of a publisher work. As one interviewee explained, *'I keep a thing called an impact and feedback register. Whenever I come across any use of our collection in a publication I enter it in that. It's a spreadsheet - it's very crude, not systematic.'* Barriers to systematically capturing citation data include lack of knowledge about how to capture this data, as well as limited staffing resources to carry out the necessary searches.⁹³

Only two UDC representatives stated that their repository uses widgets to standardise citations; in both cases, this involves a plug-in to the repository catalogue that allows users to export reference data into reference management software such as EndNote. Factors that prevent repositories from developing

⁸⁶ Interviews with project partners.

⁸⁷ Interviews with UDC representatives.

⁸⁸ Interviews with project partners.

⁸⁹ Interviews with project partners, academics and UDC representatives.

⁹⁰ Interview with publisher.

⁹¹ Interviews with academics and UDC representatives.

⁹² Interviews with UDC representatives.

⁹³ Interviews with UDC representatives.

widgets include lack of technical expertise, limited resources, and low levels of demand from users.⁹⁴

Not all UDC representatives specified whether they capture broader data on how their material is being used; of those that did provide details of data collection, the most commonly referred to indicators were number of people using the repository and use of specific items within the repository collections.⁹⁵

8.2.6 Digital Tools and Software

Both project partners interviewees emphasised that it may be more effective to focus on improving existing data capture tools and/or citation generators rather than developing a new system that relies on individuals to use. As one interviewee explained, *'what we actually need to do is develop a better means of extracting citations from existing resources and developing tools to extract multiple different types of citations, rather than persuading disparate academics to cite in the same way.'* Another interviewee observed that the current estimate of 2,500 UDC repositories in the UK amounted to a relatively small number of data points in terms of data management.⁹⁶

Interviewees across stakeholder groups noted that engaging reference management software companies to format citations to UDC materials according to a standardised model would offer another approach to overcome potential challenges around trying to change ingrained behaviour amongst academics.⁹⁷ Other interviewees noted, however, that improving reference management software would require significant investment as most software tools currently available do not produce citations to UDC materials with sufficient accuracy.⁹⁸ According to one interviewee, the feasibility of updating existing tools is likely to depend on ease with which reference management software companies can gather the data needed to rework the current platform: *'Depending on the number of individual institutions and the pages where that's presented, I don't think it's an impossible hurdle to overcome but it's something that would take time.'*⁹⁹

Project partners also raised a similar point that improving metadata on UDC repository websites was likely to be essential to working closely with citation management software companies, for example, by integrating metadata relevant for citations to UDC materials into catalogue referencing software such as Calm.¹⁰⁰ The development of repository citation widgets compatible with citation manager software was also supported more broadly by both academic and UDC representatives; as one interviewee observed, widgets that can generate citations in the standardised format automatically would be crucial to ensure support from users: *'I think that's the only way we can get this to work and be adopted by researchers.'*¹⁰¹

It should be noted, however that there remains a fairly even split between academics who prefer to create footnotes manually, and academics who use referencing management software. The main barrier to using referencing software was the time investment required to input and maintain the necessary reference data. Several interviewees noted that citations to UDC materials exported when using these tools often required further editing because the software was not able to deal with the complexity of the reference data. EndNote and Zotero were the most popular reference management software referred to by academics.¹⁰²

⁹⁴ Interviews with UDC representatives.

⁹⁵ Interviews with UDC representatives.

⁹⁶ Interviews with project partners.

⁹⁷ Interviews with project partners and academics.

⁹⁸ Interviews with project partners and academics.

⁹⁹ Interview with publisher.

¹⁰⁰ Interviews with project partners.

¹⁰¹ Interviews with UDC representatives and academics.

¹⁰² Interviews with academics.

8.2.7 Implementation

Opportunities

- 🌀 Academics perceived the primary benefit of a standardised citation model to be greater ease when citing UDC materials. Only one academic identified data collection and metrics as a potential benefit of a standardised citation model.
 - 🌀 The majority of UDC representatives were supportive of the idea of a standardised model: *'If it stops at repository code, I can't see anybody [would] object to that'*. The main benefit mentioned by interviewees was the ability to evidence the value of their collections and demonstrate impact to funders: *'there would be huge advantages in us demonstrating the use of our collections and the secondary impact'*.¹⁰³
 - 🌀 Academics felt that it would be important to coordinate implementation of a standardised model so that it happens concurrently across relevant institutions and individuals. Repositories were identified as having a key role to play in ensuring that researchers accessing their collections were aware of any standardised model.¹⁰⁴
 - 🌀 Early career researchers were identified as more likely to engage with a newly implemented model than more well-established academics.¹⁰⁵
 - 🌀 According to one publisher, greater consistency in citations would most be welcomed across the sector: *'From a publishing perspective... It would absolutely help. I can't see any negatives.'*¹⁰⁶
 - 🌀 The importance of an organisation, such as The National Archives, taking responsibility for promoting the proposed model and providing relevant information via a centralised website was highlighted by UDC representatives.¹⁰⁷
-

Challenges

- 🌀 Ensuring compliance was consistently identified as the main barrier to developing a standardised model.¹⁰⁸ A number of academics did not perceive any need to change existing practices, which they felt were sufficient for their purposes of sharing bibliographic information for UDC materials so that colleagues can follow up and consult the same sources. As one academic noted *'the last thing I want is someone to tell me how to do my footnotes'*.¹⁰⁹ The challenge around developing a consensus was also noted by one publisher.¹¹⁰
- 🌀 A number of UDC representatives also questioned whether a standardised system would be accessible for non-academic users; one interviewee said they had challenges getting users to cite them in the first place as it was seen as too academic.¹¹¹
- 🌀 Social media platforms are seen as a valuable tool for encouraging wider public engagement with UDC materials; however, as noted elsewhere, the expectations around citations on social media platforms

¹⁰³ Interviews with UDC representatives.

¹⁰⁴ Interviews with academics.

¹⁰⁵ Interviews with academics.

¹⁰⁶ Interview with publisher.

¹⁰⁷ Interviews with UDC representatives.

¹⁰⁸ Interviews with academics and UDC representatives.

¹⁰⁹ Interviews with academics.

¹¹⁰ Publisher interview.

¹¹¹ Interviews with UDC representatives.

differ considerably from traditional academic publications.¹¹²

- 🌀 Academics highlighted how the nature of archives is changing rapidly, with a move towards digitisation and online access to archive material. One interviewee noted that focusing on UDC repositories runs counter to this trend unless it can accommodate online repositories, for example, linking up the digital UDC materials with persistent object identifiers within the digital sphere.¹¹³
- 🌀 A number of academics referred to the international nature of academic practice, which they felt may impede implementation given that not all international academics accessing UDC materials held in the UK would have the knowledge or inclination to use a standardised model. One interviewee suggested taking an international approach from the outset, for example by engaging with major repositories in the UK, Europe, the US and Asia.¹¹⁴
- 🌀 Limited budget and expertise across smaller UDC repositories was seen as a potential challenge to implementing a standardised system on the ground.¹¹⁵

Implementation Partners

There were three main categories of implementation partner identified by interviewees: reference software companies such as EndNote; learned or professional societies within academic disciplines most likely to engage with UDC repositories; and publishers and funders. Opportunities to engage with reference software companies are discussed above in 'Digital Tools and Software', the latter two categories are analysed below. One project partner also highlighted the important role that Jisc would play as partners because of their digital expertise: *'it's going to need to be something suitable and adaptable for the digital age.'*¹¹⁶

All stakeholder groups felt that engaging relevant learned societies, such as the Royal Historical Society and Institute for Historical Research, to promote a standardised model would be essential to creating buy-in amongst academics.¹¹⁷ A number of interviewees suggested that these societies could play an important role in the consultation phases of developing a standardised model, as well as taking a prominent role in promoting the model during the implementation phase.¹¹⁸ As one academic observed, advocacy from learned societies may reduce the feeling that the model was *'imposed fully formed from above'*.¹¹⁹

Ensuring that publishers adopt a standardised model was identified as key strategy to ensure successful uptake, given that house style is seen as the main determinant of citation style within academic publications.¹²⁰ Interviewees noted, however, that engaging publishers and editors to adopt a standardised style may be a significant challenge; as one academic observed, *'Publishers now are quite hands off with the content of journal articles, and book publishers... are not very good at enforcing consistent academic practice.'*¹²¹ Low levels of engagement from publishers was also evident in the recruitment process for the current phase of stakeholder interviews, further indicating potential challenges in developing partnership working with this key sector.

At the same time, interviewees highlighted the strategic benefit for publishers of supporting a standardised model. Potential benefits for publishers include data harvesting and generating income from click-through

¹¹² Interviews with academics and project partners.

¹¹³ Interviews with academics.

¹¹⁴ Interviews with academics.

¹¹⁵ Interviews with UDC representatives.

¹¹⁶ Interviews with project partners.

¹¹⁷ Interviews with academics, UDC representatives, project partners and publisher.

¹¹⁸ Interviews with academics and project partners.

¹¹⁹ Interviews with academics.

¹²⁰ Interviews with academics, UDC representatives, and project partners.

¹²¹ Interviews with academics.

links between repositories webpages and e-publications.¹²² Strategies for engaging publishers mentioned by interviewees include inviting publishers to working groups as part of the consultation phase; developing a centralised webpage or database setting out the requirements of a standardised model so that publishers could simply point authors towards that centralised guidance; and reaching publishers through industry conferences during implementation phase.¹²³ As one publisher noted, *'we would want to have a certain element of confidence that it would reach all tipping points that people will recognise and use that option.'*¹²⁴

Research councils were also seen as a key sector to engage to ensure the successful implantation of the proposed model, for example, by making use of a standardised model a condition of funding.¹²⁵ According to one interviewee, *'the research councils need to do more in insisting to the adherence to the highest standards of academic practice in publication.'*¹²⁶ A number of interviewees noted that this may be difficult to enforce in practice and emphasised that recommendations rather than requirements would be preferable for both funders and practitioners. This was echoed in feedback from one of the main research councils: *'If something existed, we could certainly have something on our website. It's not that it's not a priority for us - just that our guidance doesn't go that far.'*¹²⁷

¹²² Interviews with UDC representatives.

¹²³ Interviews with academics and publisher.

¹²⁴ Interview with publisher.

¹²⁵ Interviews with academics, UDC representatives, and project partners.

¹²⁶ Interviews with UDC representatives.

¹²⁷ Email feedback from a research council following interview request.

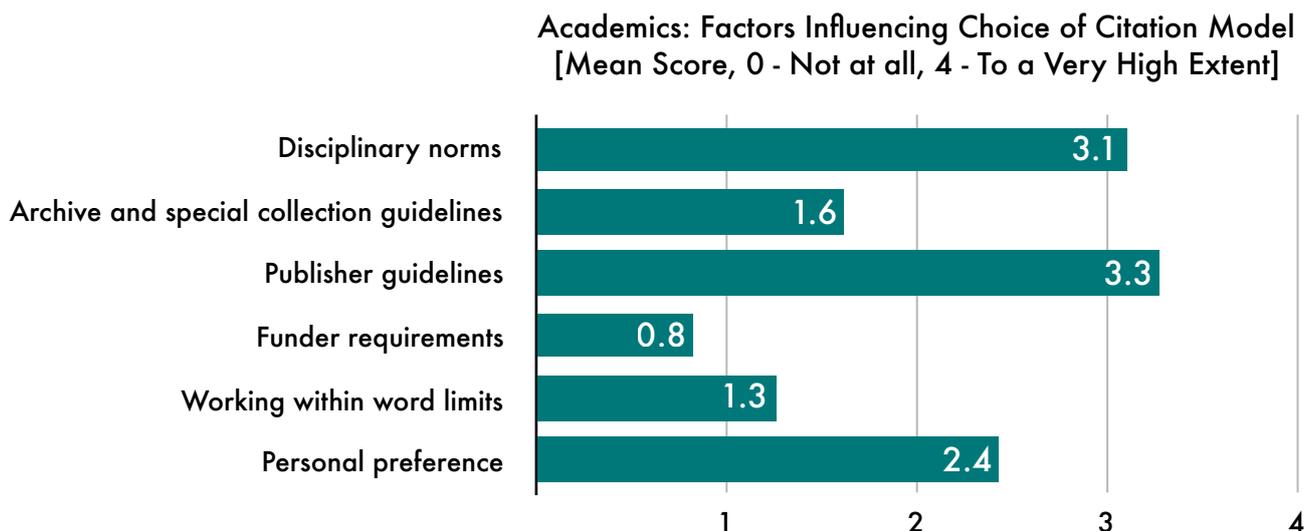
8.3 Survey Analysis

8.3.1 Citation Model

Academic Citation Practices

Chicago was the most commonly used citation model amongst the academics surveyed (54% of participants). A small proportion of survey participants used other citation models, such as Harvard (14%), MLA (7%) and APA (2%). Almost a quarter (23%) said they used other models, of which six said they followed a style depending on the publisher, five said they used MHRA and two used their institution's internal style.

The three main factors influencing academics' choice of citation model included publisher guidelines (mean score of 3.3 of 4), disciplinary norms (score of 3.1) and personal preferences (score of 2.4). UDC repository guidelines did not appear to be considered very influential among the academics surveyed, with an average score of 1.6.



When asked if they followed any particular style for citing UDC collections, the most common answer among the academics surveyed (26 individuals) was that they used the repository catalogue reference. Other survey participants gave answers that were similar to the above questions and said they followed the repository's style (14 individuals), publishers' styles (10 individuals), their own institution's style (three individuals) or disciplinary norms (two individuals). Seven survey participants said they did not follow any particular style for UDC citations.

Use of Abbreviations

43% of repositories surveyed have an established and recognised abbreviation for their repository; of these, 61% believe that this abbreviation is unique to their repository, 21% believe that it is not unique and 18% are not sure whether it is unique. Of those who specified the code type used, 33 used a text abbreviation code, two used a combination of text/numbers (ARCHON) and three repositories did something else (e.g. a combination of text abbreviation with full words). Of the 33 repositories which used a text only code, 23

used a three letter code, seven used a four letter code and five used a five or more letter code .¹²⁸

71% of academics surveyed said that they use abbreviations when referencing archives and special collection repositories; the remaining 29% said that they did not. 31% of the abbreviation examples given by academics were two letter codes (e.g. BL); 42% were three letter codes (e.g. TNA); 18% were four letter codes (e.g. NARA); and 9% were five or more letter codes (e.g. PRONI).

Level of Variation in UDC Repository Citations

Academics were asked to rate the level of variation with regards to how archive and special collection repository names are cited in publications, while repository representatives were asked to rate the level of variation for their own repository name. The results indicate a moderate degree of variation in terms of how UDC repositories are cited, with a mean score of 2.0 across all the survey participants. There was no statistically significant difference in views between repository staff and academics.

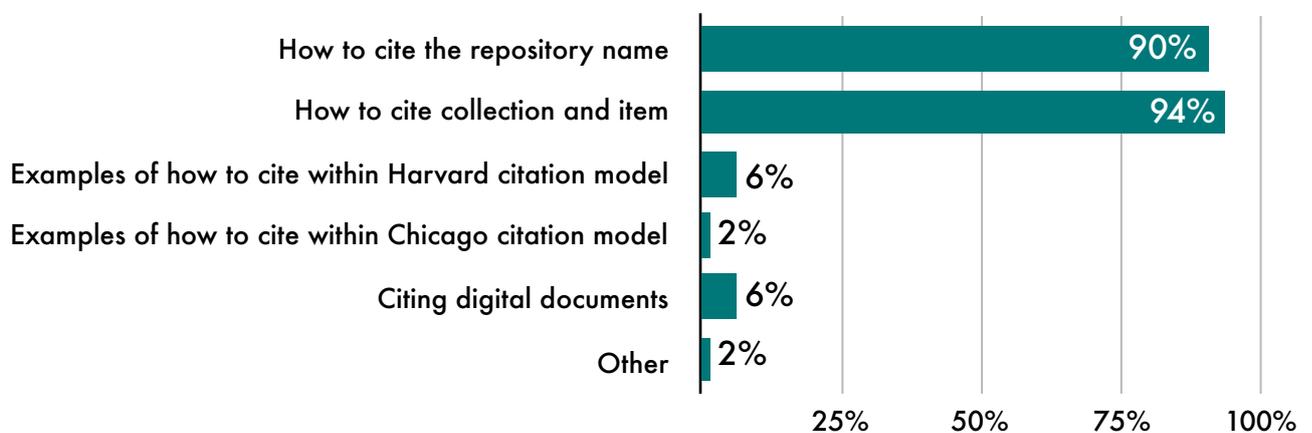
Level of Variation in UDC Repository Citations
[Mean Score, 0 - No Variation, 4 - Very High Degree of Variation]



8.3.2 Current Practice

Repository Citation Guidelines

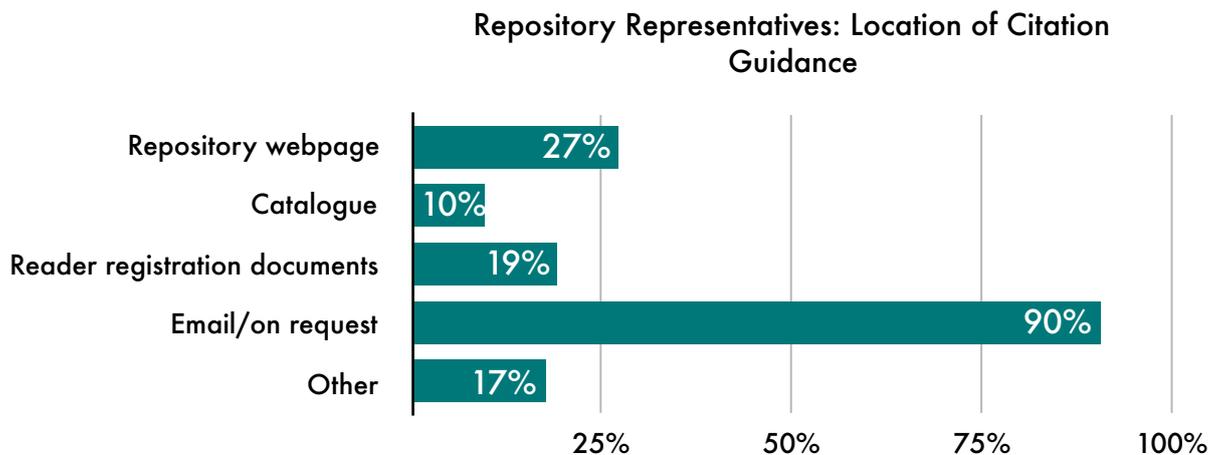
Repository Representatives: Guidelines for Users on Citing Repositories in Academic Publications



¹²⁸ Note that some repositories used more than one text abbreviation type/length, hence the sub-categories adding up to more than 33.

The majority (70%) of repository representatives surveyed said they provide guidance for users on how to cite their repository within academic publications, of which almost all provide guidance around how to cite the collection and item (94%) and how to cite the repository name (90%). A very small proportion of the repository representatives said they provide guidance on how to cite digital documents (6%) and how to cite within Harvard (6%) or Chicago (2%) citation models.

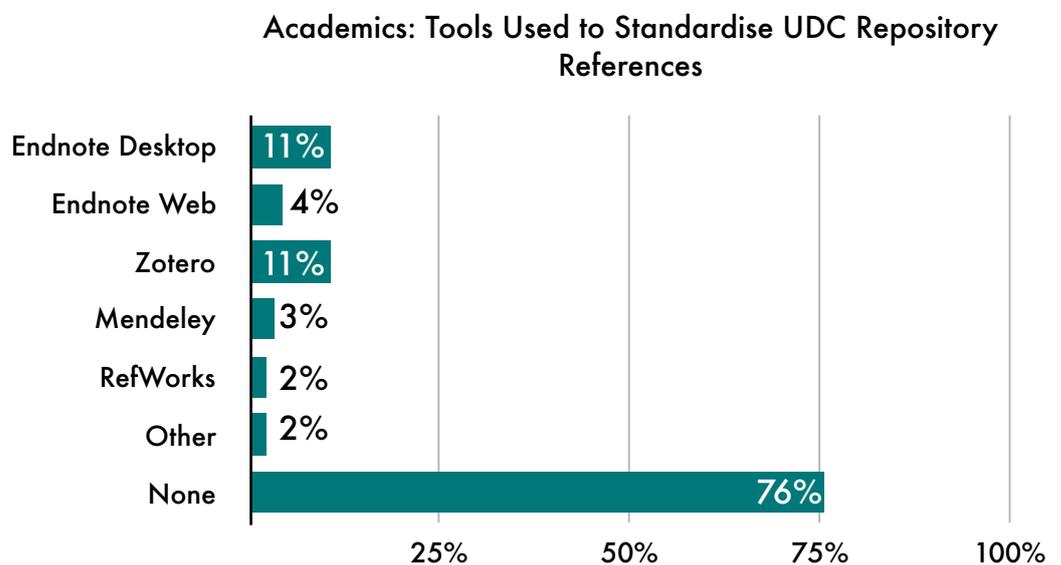
Citation guidelines do not appear to be widely accessible to users in many UDC repositories; 90% of the repository representatives surveyed said they provided citation guidance via email or on request, whereas less than a third (27%) provided guidance on the repository webpage. Among participants that selected other options (17%), five said they provide guidelines through leaflets, forms or permission letters, while another two said they gave guidance verbally.



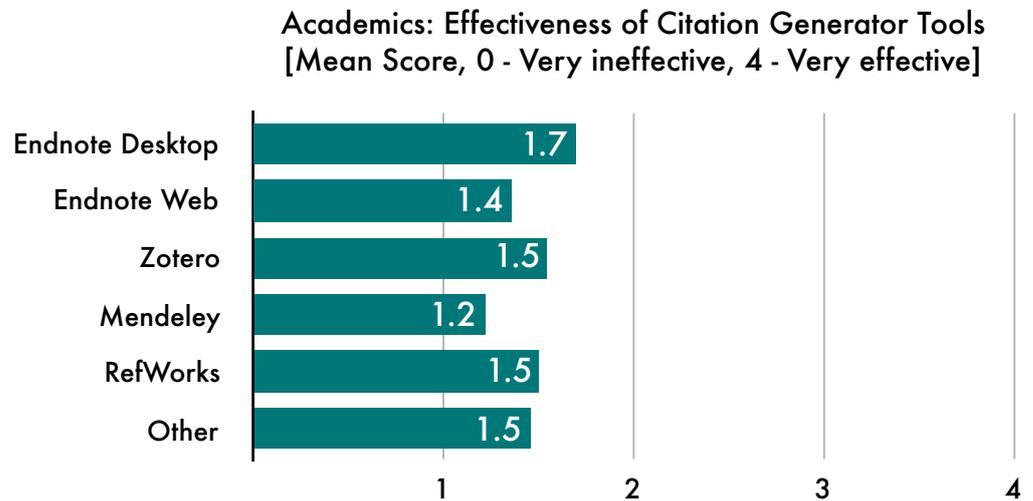
Only 8% of repository representatives surveyed said they provided other support for users, such as widgets or other citation generators, to standardise references to their collections, while 85% did not and 6% did not know.

Use of Widgets and Citation Generators

Three quarters (76%) of the academics surveyed did not use widgets or other citations to standardise references to UDC repositories in their publications. Only a small proportion said they used specific tools, such as Endnote Desktop (11%) or Zotero (11%).



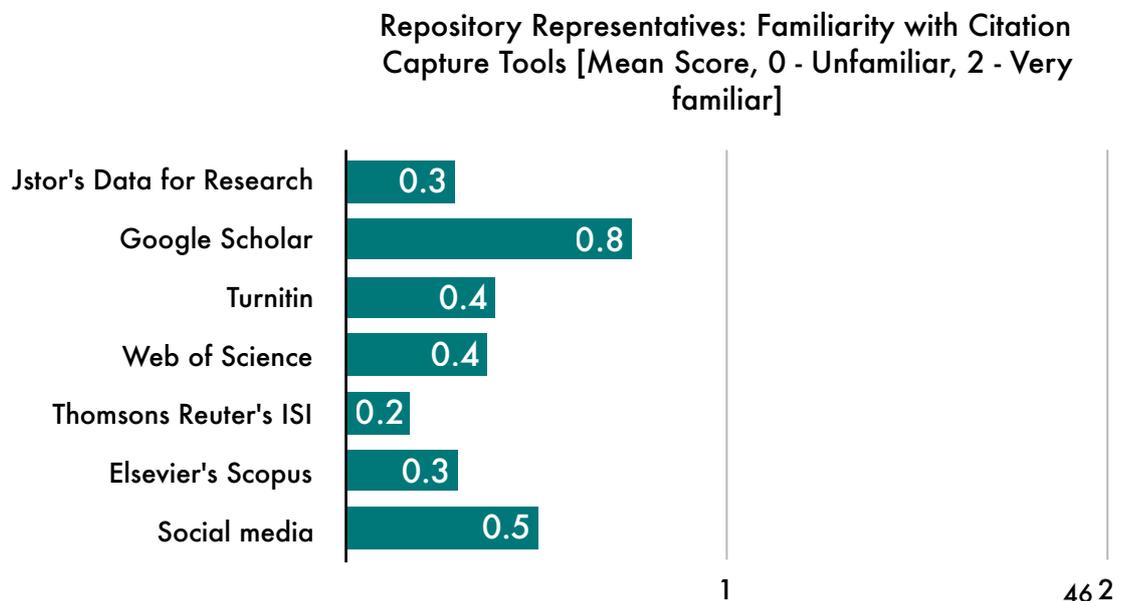
Citation generator tools were not seen as particularly effective among the academics surveyed. Among the survey participants that used citation generators, Endnote Desktop was rated as being the most effective, with a mean score of 1.7 of a highest possible score of 4. Mendeley was rated as being the least effective citation generator tool, with a mean score of 1.2.



Approaches to Demonstrate Use of UDC Repository Material

The majority of repository representatives surveyed said they tracked the type of repository users (80%) and the purpose of their use of the collections (69%). Among the repositories that captured this information, 76% did so through registration data. Just under half the participants (47%) used sign-in books, 28% used ad hoc user surveys and 26% asked verbally.

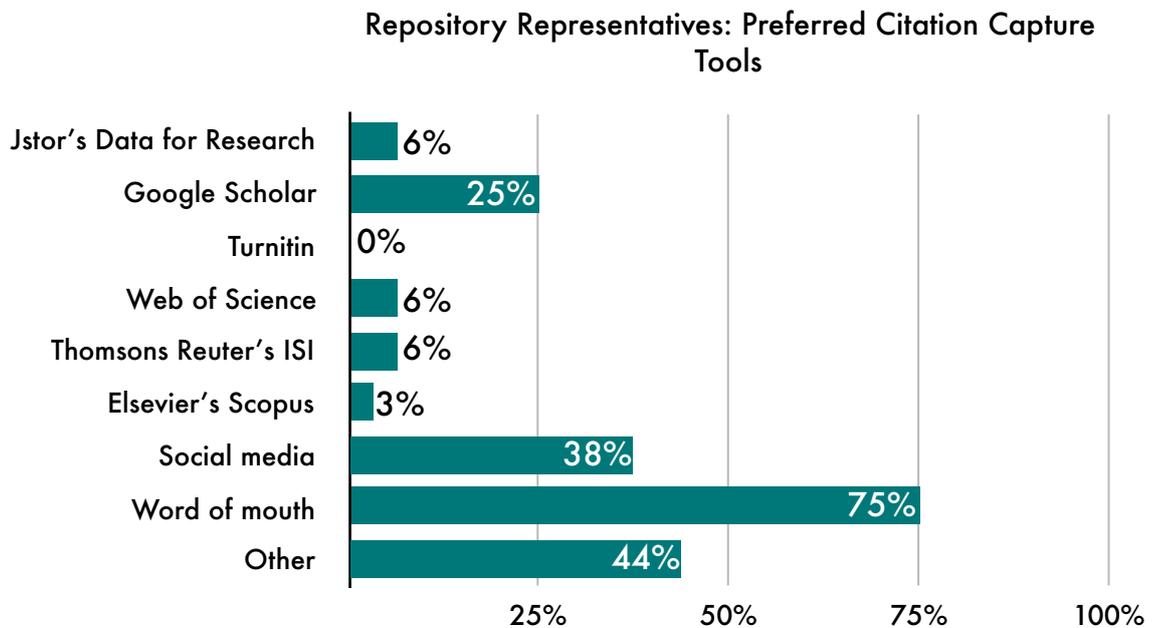
Survey responses indicate that UDC repositories may not be making the most of capturing citations to their collections as a result of a lack of awareness of existing citation capture tools. While repository representatives appeared to be somewhat familiar with Google Scholar (average score of 0.8), they did not appear to be very familiar with other tools, especially Thomsons Reuter's ISI (score of 0.2), Jstor's Data for Research (score of 0.3) and Elsevier's Scopus (score of 0.3).



More than half the repository representatives (57%) said they did not collect data on published citations to their UDC collections; another 7% planned to do so in the future. Another 7% of repository representatives surveyed said they collected data on citations, while 30% said they did so partially.

The main reasons offered by the repository representatives for not capturing published citations were a lack of resources¹²⁹, including staff and time, as well as it not being considered a priority.¹³⁰ Some participants said that the number of citations to their repository was too low for citation capture to be relevant, while others said that it was not something that they were required to report on. Nine repository staff said they did not capture citations due to a lack of relevant tools or knowledge in how to do so, while five found the scale of such a task to be daunting. Three survey participants did not appear to see any need for making any explicit efforts to capture citations, as they said their users would keep them informed of any relevant publications involving their collections.

Among the survey participants that already captured citations, 61% captured primary or original citations, 6% captured secondary citations and 33% captured both. Three quarters (75%) of the repository representatives surveyed said word of mouth was the most preferred method for finding published citations to their UDC collections, while a minority preferred using social media (38%) or Google Scholar (25%).



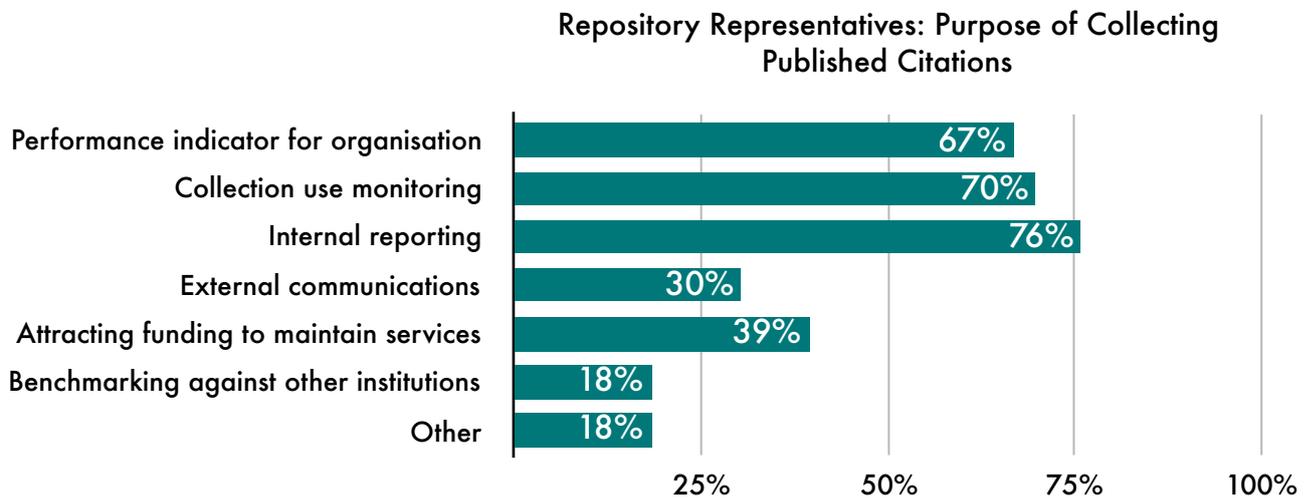
More than two fifths (44%) of survey participants preferred other methods or tools, of which eight individuals captured citations directly from books or other published material. Some survey participants identified books through checking relevant sites like Amazon or through cross referencing repository visits, while others asked visitors to send them published material directly or gathered this information from user registration or permission forms. Less common methods used for capturing citations included monitoring Wikipedia citations, internal references, alerts with relevant journals or automatic tracking through Digital Object Identifiers (DOIs).

The top three reasons for collecting published citations mentioned by the repository representatives surveyed included internal reporting (76% of participants), monitoring use of collections (70%) and as a performance indicator for the organisation (67%). A minority of the survey participants said they collected published citations in order to attract funding to maintain services (39%), use in external communications (30%) or for benchmarking against other institutions (18%).

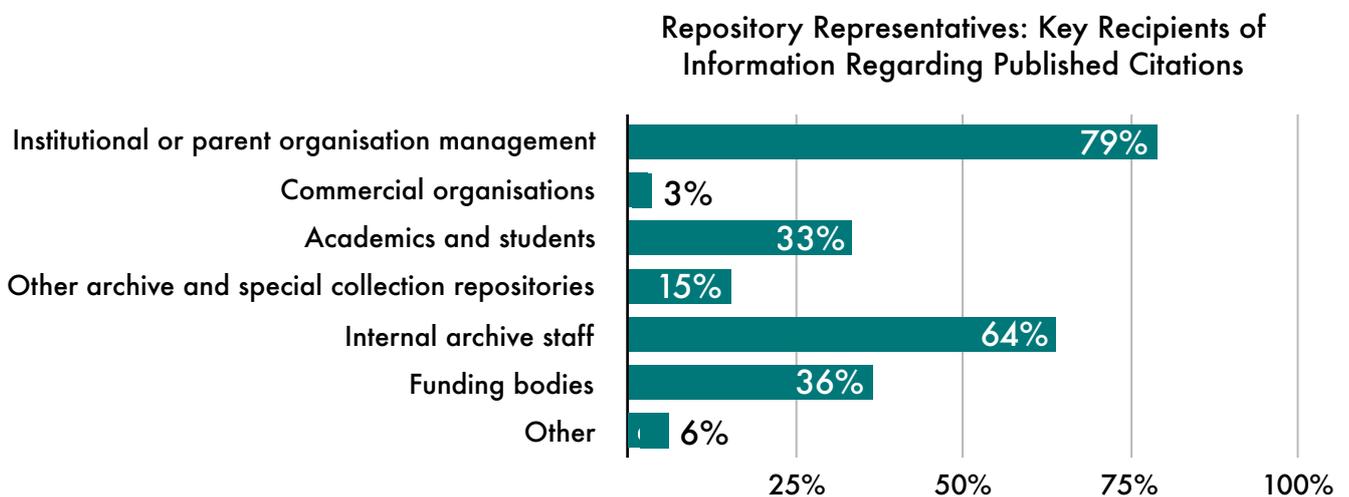
¹²⁹ 17 participants.

¹³⁰ 17 participants.

Among individuals identifying other reasons, three said they collected citations in order to provide information about collections or published works involving collection items to users; other reasons offered included internal advocacy, interest in the use of collections or search engine optimisation (SEO).



The main recipients of information regarding citations to repositories included institutional or parent organisation management (79% of participants) and internal archive staff (64%). Around a third of survey participants said key recipients included funding bodies (36%) or academics and students (33%), while 15% said other archive and special collection repositories. Only 6% collected information on published citations for commercial organisations.

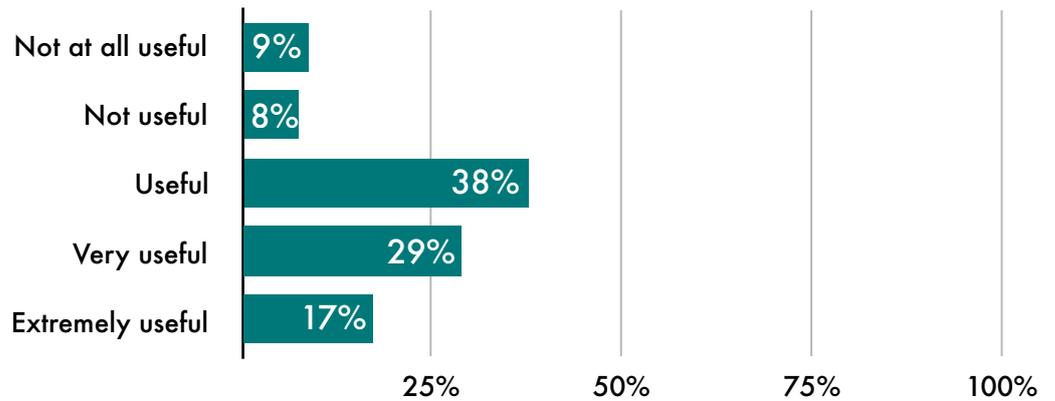


8.3.3 Opportunities and Challenges

Perceived Usefulness of Standardised Abbreviations

Survey responses suggest that there is a high level of support among academics for a system of standardised abbreviations for UDC repositories; a significant majority (84%) of academics surveyed thought a system of standardised abbreviations would be useful, very useful or extremely useful.

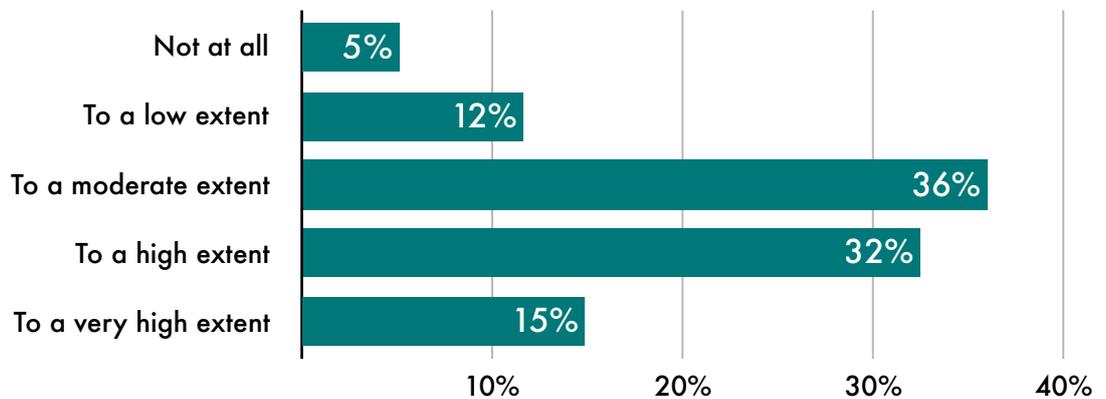
Academics: Usefulness of Standardised Abbreviations



Extent of Support for Standardised Abbreviations

The graph below combines repository and academic views on the expected support among all groups (repositories, academics and publishers/funders) for a system of standardised abbreviations for archive and special collection repositories. 36% of participants thought that there would be moderate support across user groups, and 32% thought this support would be high.

Expected Support for Standardised Abbreviations for UDC Repositories [All User Groups]



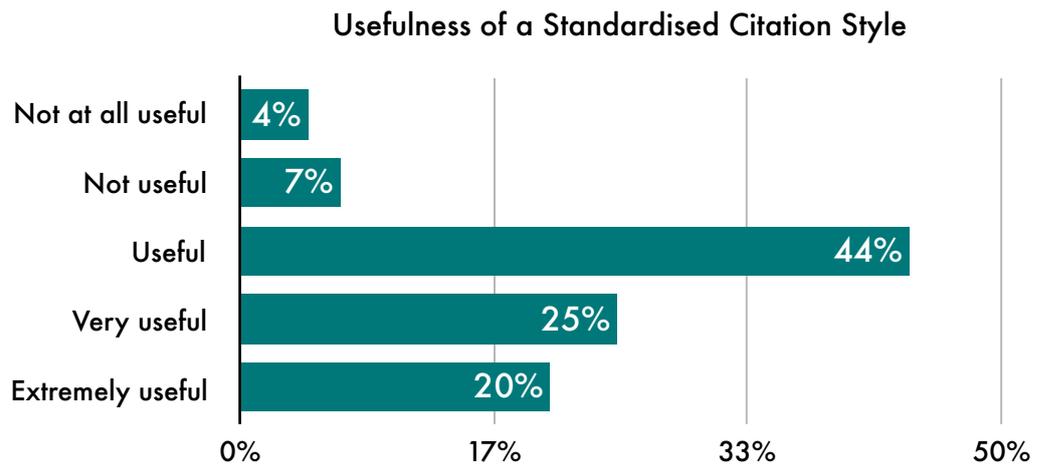
These views represent a mean score of 2.4 out of a possible maximum of 4 (where '0' represents 'not at all' and '4' represents 'to a very high extent').

Support was anticipated to be greatest among repositories, with a mean score of 2.6, then among academics, with a mean score of 2.4, and lastly funders and publishers, with a mean score of 2.1. There was no statistically significant difference in the anticipated levels of report for each of these groups among the repository representatives and academics surveyed.¹³¹

¹³¹ The mean score for the anticipated level of support among academics was higher for the repository staff surveyed (mean score of 2.6) than for the academics (mean score of 2.3), but this was not a statistically significant difference.

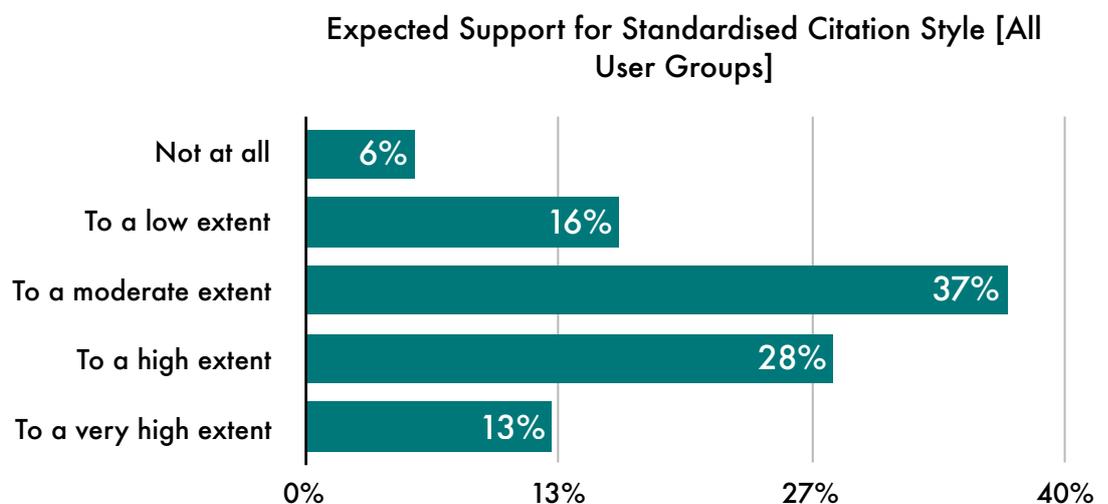
Perceived Usefulness of a Standardised Citation Style

Survey participants were asked how useful they thought a standardised archive and special collections citation style would be either to their repository (repository representatives) or to themselves and other researchers (academics and researchers). The vast majority (89%) of survey participants thought that it would be useful, very useful or extremely useful. There was no statistically significant difference in the perceived level of usefulness among the repository staff and academics surveyed.¹³²



Extent of Support for Standardised Citation Style

The graph below combines repository and academic views on the expected support for a standardised citation style among all user groups, including repositories, funders, publishers, academic and other users. 37% of survey participants thought that there would be moderate levels of support, while 41% thought that the level of support would be high or very high.



The mean score for the expected level of support among all the user groups was 2.3 of a maximum possible

¹³² The mean score for the level of usefulness was 2.6 out of a possible 4 for repository staff and 2.4 for academics, but this was not a statistically significant difference.

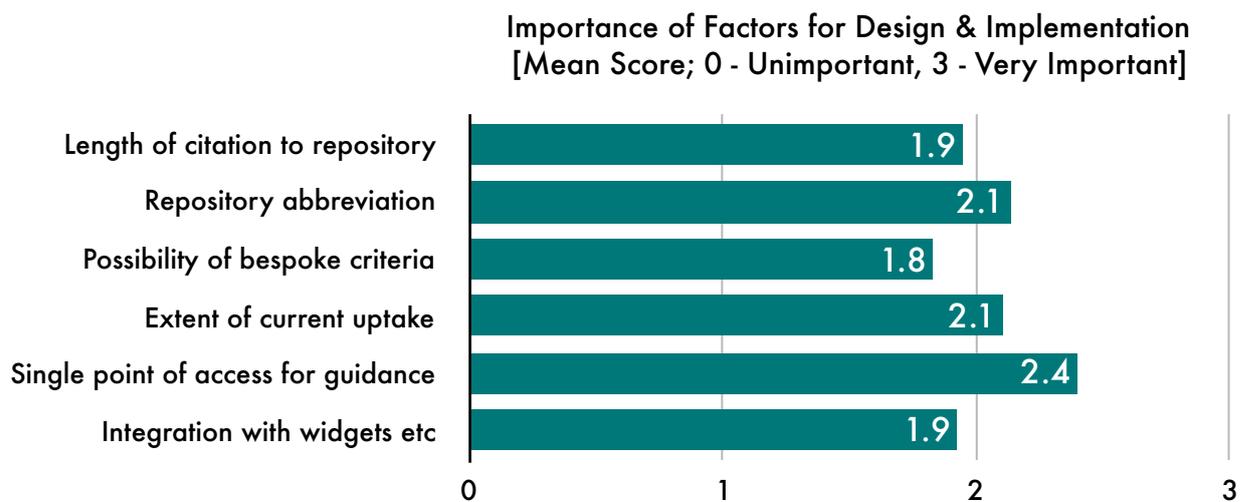
score of 4, with '0' representing 'not at all' and '4' representing 'to a very high extent'. The level of support was expected to be highest among repositories (mean score of 2.7), followed by academics (mean score of 2.6). Survey participants also thought there would be moderate levels of support among publishers and funders, with mean scores of 2.2 and 2.0, respectively. Repository staff and academics expected other users to be the least supportive of a standardised citation style, with a mean score of 1.7.

Repository representatives anticipated the level of support among academics to be slightly greater than academics did themselves, with mean scores of 2.6 and 2.5, respectively. There were no other statistically significant differences in views between repository representatives and academics regarding the expected level of support for a standardised citation style.¹³³

8.3.4 Implementation

Designing and Implementing a Standardised Citation Style

Survey participants were asked how important various factors would be for designing and implementing a standardised citation style. Having a single point of access for citation guidance was thought to be the most important factor (mean score of 2.4), followed by repository abbreviation (mean score of 2.1) and the extent of current uptake for the style (mean score of 2.1).



Repository staff considered some factors to be more important compared with academics. The average score for the importance of integrating widgets and other citation generators was 2.1 among repository representatives and 1.8 among academics. In terms of the importance of a single point of access for citation guidance, the score among repository staff was 2.5 while the score for academics was 2.3. There were no other statistically significant differences in views between repository representatives and academics.¹³⁴

¹³³ Differences between repository staff and academics which were not found to be statistically significant included the expected level of support among repositories, funders, publishers and other users. The mean score for these groups varied from 0.1 to 0.2 points between repository staff and academics.

¹³⁴ Differences between repository staff and academics which were not found to be statistically significant included the importance of the length of citation to repository, repository abbreviation and the possibility of bespoke criteria for specific repositories. There was a 0.1 point difference in the mean score among repository staff and academics in these areas.

8.3.5 Survey Data Tables

Survey Characteristics

		Percentage	Number
Total Number of Participants		100%	184
Participant Group	Repositories	49%	90
	Publishers and funders	0%	0
	Academics	51%	94
Job Role	Archivist	33%	60
	Librarian	10%	19
	Other-heritage/information professional	6%	11
	Publisher/Editor	0%	0
	Funding Manager	0%	0
	Academic	35%	64
	Other Researcher	16%	30

Repository Staff

			Percentage	Number
Type of Repository	Local Authority		17%	15
	Business		2%	2
	Private		6%	5
	Specialist		6%	5
	Higher Education		46%	41
	Museum or Cultural Institution		13%	12
	Other		10%	9
Size of Repository Holdings	Average Printed Material	Cubic Metres	-	1,377
		Linear Metres	-	27,623
	Average Digital Material (Gigabytes)		-	38,923
Repository Head Count	Less than 10		58%	45
	10-29		23%	18
	30-99		12%	9

		Percentage	Number
	100 and above	8%	6
Repository Material	Printed volumes	86%	77
	Archives and manuscripts	89%	80
	Cartographic materials	57%	51
	Visual and audiovisual materials	82%	74
	Born-digital materials	66%	59
	Research data	32%	29
	Other	21%	19

Academics and Other Researchers

		Percentage	Number
Subject Area	Agriculture and related sciences	0%	0
	Architecture, building and planning	1%	1
	Biological sciences	0%	0
	Business, management and administrative studies	1%	1
	Computer science	1%	1
	Creative arts and design	2%	2
	Education	2%	2
	Engineering and technology	0%	0
	English	11%	10
	Historical and philosophical studies	67%	63
	Languages	0%	0
	Law	1%	1
	Mathematical sciences	0%	0
	Medicine and allied subjects	0%	0
	Music	3%	3
	Physical sciences	2%	2
	Psychology	0%	0
Publishing, media and information management	0%	0	

		Percentage	Number
	Social studies	1%	1
	Sport	0%	0
	Veterinary science	0%	0
	Other	7%	7
Frequency of Using Archive and Special Collection Materials in Their Work	Never	0%	0
	Rarely	7%	1
	Sometimes	16%	0
	Very often	78%	1

Academics and Other Researchers: When did you last do any of the following? (Percentage)

	Never	More than a year ago	In the last year	In the last six months	In the last month	In the last week
Visit an archive or special collection repository	1%	6%	4%	19%	44%	25%
Access digitised special collection material online	2%	2%	2%	11%	22%	61%
Use archive or special collection material in your work	1%	2%	3%	5%	18%	70%
Access born-digital materials online	20%	5%	13%	9%	22%	31%

Academics and Other Researchers: When did you last do any of the following? (Number)

	Never	More than a year ago	In the last year	In the last six months	In the last month	In the last week
Visit an archive or special collection repository	1	6	4	18	41	23
Access digitised special collection material online	2	2	2	10	21	57
Use archive or special collection material in your work	1	2	3	5	17	65
Access born-digital materials online	19	5	12	8	20	29