

Closing Gaps: A Model of Cumulative Curation and Preservation Levels for Trustworthy Digital Repositories

Jonas Recker
GESIS – Leibniz Institute for the
Social Sciences

Mari Kleemola
Finnish Social Science Data Archive,
Tampere University

Hervé L'Hours
UK Data Archive, UK Data Service,
University of Essex

Abstract

Curation and preservation measures carried out by digital repository staff are an important building block in maintaining the accessibility and usability of digital resources over time. The measures adequate to achieve long-term usability for a given audience strongly depend on scenarios of (re)use, the (intended) users' needs and skills, the organisational setting (e.g., mission, resources, policies), as well as the characteristics of the digital objects to be preserved. The assessment of curation and preservation measures also forms an important part of existing certification procedures for trustworthy digital repositories (TDRs) as offered, for example, by the CoreTrustSeal foundation, the nestor network, or ISO.

The digital curation community is presented with the challenge of finding community-, organisation-, and object-specific approaches to curation and preservation at the same time as defining the minimum level of curation and preservation measures expected from a TDR in sufficiently generic terms to ensure applicability to a wide array of repositories. Against this backdrop, this paper discusses the need for and benefits of community-agreed levels of curation and preservation to address this challenge, and considers the tiered model proposed by the CoreTrustSeal Board as an example.

The proposed model is then applied in an analysis of successful CoreTrustSeal applications from 2018–2022 in an effort to better understand the capacity of the curation and preservation levels to capture the respective practices of repositories and to identify potential gaps.

Submitted 9 February 2024 ~ *Accepted* 22 February 2024

Correspondence should be addressed to Jonas Recker, GESIS – Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Cologne, Germany. Email: jonas.recker@gesis.org

This paper was presented at the International Digital Curation Conference IDCC24, 19-21 February 2024

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introduction

Curation measures carried out at ingest and over the lifespan of digital objects held in and disseminated by repositories are crucial in maintaining access to, understanding, and continued use of digital assets. Such measures may consist in creating or enriching metadata, documentation, and other context information (e.g., Representation Information and Preservation Description Information as defined in the OAIS Reference Model (Consultative Committee for Space Data Systems, 2012)), creating copies in archival and/or dissemination formats, correcting or enhancing the content of the objects, putting them in emulation environments suitable for allowing them to be used, or performing file format migrations to prevent technological obsolescence. Models such as the DCC Curation Lifecycle model (Higgins, 2008) or the DCN's CURATE(D) steps (Data Curation Network, 2022; Johnston et al., 2018), catalogues such as the American Research Libraries Association's *SPEC Kit 354* (Hudson-Vitale et al., 2017), and metadata schemas such as PREMIS (PREMIS Editorial Committee, 2015) list and describe up to 50 different actions considered relevant to the curation and preservation of digital assets.

Which specific measures are considered adequate and necessary to achieve the objective of continued accessibility and usability, and which measures are in fact performed, depends on numerous factors. These may be specific to the domain and organisational framework in which the digital objects are preserved and disseminated (including policies, available resources, and expertise), the audience for whose use the digital assets are curated, or to the characteristics of the digital assets themselves. As a result, clearly one size does not fit all when it comes to determining which curation and preservation measures are considered adequate to enable and maintain access and re-use over time for a given collection and community.

This presents challenges to repositories, e.g., when seeking to design and implement a robust curation and preservation process, but also to the assessment, audit and certification efforts for trustworthy digital repositories (TDRs) offered, for example, by the CoreTrustSeal Foundation (CoreTrustSeal Standards and Certification Board, 2022), the nector network (nector Certification Working Group, 2013), and the International Organization for Standardization (International Organization for Standardization, 2012). Certification entails assessing compliance with a standard which must be sufficiently generic to allow for the assessment of a broad array of repositories but specific enough to help determine whether the curation and preservation measures performed are fit to achieve the goals the repository has with regard to long-term access and use of its digital collection.

To address this issue, the CoreTrustSeal Board proposed a model of cumulative curation and preservation levels and revised it based on input from the community. The model is intended as an instrument supporting repositories in the process of systematically planning their curation and preservation efforts as well as facilitating TDR (self-)assessment and certification. The proposed levels provide a common basis for the wider curation and preservation community to discuss and agree on what minimum requirements for TDRs are and, by extension, facilitate in-scope/out-of-scope decisions in TDR certification.

Curation and Preservation Levels for Trustworthy Digital Repositories

The concept of TDRs, initially short for Trusted Digital Repositories, but today commonly used to refer to *Trustworthy* Digital Repositories, emerged along with digital repositories and evolved as digital preservation took shape as a professional discipline.

Definitions of TDRs tend to highlight the following characteristics: TDRs have

- an explicit mission and responsibility to preserve and provide long-term access to a collection of digital assets;
- a defined community for whose access and use digital assets are preserved;
- appropriate infrastructure, including organisational and technical systems in accordance with agreed standards, adequate to achieving the mission and operated transparently;
- organisational, including financial, sustainability;
- sufficient expertise and capacity to take actions to mitigate risks associated with technological and organisational change as well as changes in the needs and skills of the user community. (See CoreTrustSeal Standards and Certification Board, 2022; International Organization for Standardization, 2012; nestor Certification Working Group, 2013; Research Libraries Group, 2002).

A need for an active management of digital assets and their metadata, including curation and preservation measures, is implied in the listed characteristics, and spelled out in the standards supporting the certification and audit processes for TDRs cited here. Each of these standards contains criteria relating to the digital curation lifecycle steps as defined by the Digital Curation Centre:

- Conceptualising and creating information packages (data and metadata).
- Making provisions for access and use.
- Appraising and selecting digital objects in need of and suitable for preservation.
- Reappraising and disposing of digital objects.
- Ingesting digital objects into the repository and storing them securely.
- Undertaking preservation actions and transforming digital objects when required.
- Ensuring continued access and re-usability. (See Higgins, 2008).

The What and Why of Levels of Curation

The lists and descriptions of activities mentioned above (and others like them) define the spectrum of activities that the community considers as relevant in the process of curating and preserving digital objects. They help answer the question of what should be done to ensure continued access, use, and understanding of the digital assets. However, they do not provide an answer to the question of how these actions should be performed, and which specific combinations and implementations of actions constitute the minimum that needs to be done to reach a given goal. This requires considering relevant recommendations and standards together with characteristics of the digital assets, the user community's needs and knowledge base, as well as organisation-specific conditions and policies.

For example, while there is wide consensus that three redundant copies of digital assets on different types of storage media and in different locations constitute good back-up practice, repositories working with sensitive data or dealing with high volumes of data may find different approaches to back-up more adequate in their specific situation as they have to weigh the risk of data loss against the risk of information breaches or the cost for storage. And even though file formats with open specifications are generally recommended to facilitate long-term preservation and re-use, they may not be a feasible option, e.g., when data comes from instruments providing their output only in a proprietary, manufacturer-specific format, or when considerable parts of

the designated community predominantly use proprietary software for data processing and analysis, as is the case in the quantitative social sciences.

The need for community-, organisation- and object-specific considerations and solutions poses a challenge for repositories when tasked with developing curation and preservation policies and processes, for those who want to communicate their curation and preservation practices to their stakeholders, as well as for (self-) assessment and certification efforts. Common certification standards and requirements must be sufficiently generic enough to apply to a wide range of potential services, but specific enough to be able to differentiate between and assess different curation and preservation approaches.

Community-agreed, concise levels of curation and preservation are an instrument which helps in closing these gaps. They can “provide an important reference point for digital objects’ depositors, funders and (re)users and for collaborations and partnerships communally offering (meta)data services” (Recker, L’Hours, & Kleemola, 2023). For example, the digital repository staff at University of North Carolina Chapel Hill and Duke University developed and applied a set of curation levels based on the *DCN CURATE(D)* and *SPEC Kit 354* curation activities (Data Curation Network, 2022; Hudson–Vitale et al., 2017), finding that the levels

‘facilitated communication and transparency within and between our institutions. Each institution considered and implemented strategies to communicate and make transparent the curation activities provided, why certain tasks are important, and what value they add to a diverse set of stakeholders [...]. We found that while “curation” may still be an obscure term to a broader University audience, being specific about the activities performed helped to create more transparency around our work.’ (Lafferty–Hess et al., 2020, pp. 10–11)

Similarly, the ability to reference clearly defined levels of curation and preservation supports the provision of object-level information about the amount of care a given object has been or is receiving as part of this object’s Preservation Description Information (Consultative Committee for Space Data Systems, 2012, Chapter 4.2.1.4.2), allowing repositories to communicate this information both to human users or to machines processing the objects and/or their metadata. This can greatly support efforts to increase (meta)data interoperability, e.g., in the context of ‘FAIRification’ (see, for example, Welter et al., 2023).

For TDR assessment and certification efforts, tiered models defining distinct levels of curation and preservation can provide a framework for applicants to systematically plan and describe their curation activities, as well as facilitating decision-making on whether a given service is in scope for TDR certification. The proposed CoreTrustSeal Curation and Preservation Levels (CoreTrustSeal Standards and Certification Board, 2023) support both scenarios.

The Proposed CoreTrustSeal Curation and Preservation Levels

In contrast to the nestor and ISO standards, the CoreTrustSeal Requirements have always mandated applicants to state the extent of curation that submissions to the repository undergo in the form of curation levels (see Table 1). The Requirements state that this initial curation, happening once as digital objects are deposited and ingested into the repository, must be complemented with digital preservation measures, assuming that

‘(1) initial deposits are retained unchanged and that edits are only made on copies of those originals, (2) metadata that enables the Designated Community to understand and use the data independently (i.e., without having to consult the original creator) is present at deposit or added by the repository, and (3) ongoing measures for active preservation are in place for the greater part of the collection(s).’ (CoreTrustSeal Standards and Certification Board, 2022, p. 7)

However, while providing valuable context information for the reviewers and future readers of the successful public applications, the levels as they currently stand do not accurately represent the realities of curation practice, nor are they a reliable means to determine whether an applicant is in scope for CoreTrustSeal certification. The current levels do not make it sufficiently clear whether they refer to curation at ingest or if responses are meant to cover measures of curation over time (as is suggested, among other things, by the reference to 'active preservation' in the guidance quoted above). Just as the relevance of 'active preservation' is implied, but not made explicit in the guidance, the existing levels do not allow CoreTrustSeal reviewers to infer whether an applicant is in scope for CoreTrustSeal, or not.

The current levels are also not a comprehensive representation of the breadth of curation practices performed at the point of deposit and ingest. For example, they do not allow applicants to systematically capture the extent of practices of (meta)data checking, technical interventions such as file format conversions ('normalisation') or the creation of hard-/software environments in which the digital objects can be rendered by the users, nor do they adequately capture scenarios in which extensive checks are carried out during ingest, where documentation and metadata are entirely created by repository staff, or where data harmonisation is carried out during ingest resulting in a new data object, going far beyond a simple 'editing of deposited data [for accuracy]'.¹

Table 1. Current CoreTrustSeal Levels of Curation (CoreTrustSeal Standards and Certification Board, 2022, para. R0).

Level	Description
A.	Content distributed as deposited
B.	Basic curation – e.g., brief checking, addition of basic metadata or documentation
C.	Enhanced curation – e.g., conversion to new formats during ingest, enhancement of documentation and metadata
D.	Data-level curation – as in C above, but with additional editing of deposited data

In response to these issues, the CoreTrustSeal Board has proposed an alternative model of Curation and Preservation Levels and revised it based on comments and suggestions from the community (CoreTrustSeal Standards and Certification Board, 2023; Recker, L'Hours, & Kleemola, 2023). The proposed levels (see Table 2) distinguish different degrees of engagement with submitted content upon receipt (from 'none' to 'initial curation') as well as two levels of preservation, focusing on the preservation measures for (technical) long-term usability and understandability respectively. Thus, the model allows repositories to express how they engage with the digital objects submitted to and ingested into their holdings both during ingest and over time. The initial curation-focused levels D. and C. can be freely combined with the long-term preservation-focused levels B. and A. to express the level of care that applies to a given collection of digital objects and accordingly, each digital object in that collection. The levels defined in the proposed model thereby facilitate the provision of information about curation and preservation measures in the form of repository, catalogue and object-level metadata.

¹ Note that the Requirements versions 2017–2019 and 2020–2022 contained the following phrasing for Level D: 'D. Data-level curation – as in C above, but with additional editing of deposited data for accuracy' (see CoreTrustSeal Standards and Certification Board, 2019).

Table 2. Proposed CoreTrustSeal Curation and Preservation Levels (CoreTrustSeal Standards and Certification Board, 2023).

Level	Description
Z. Zero.	Content distributed as deposited. Unattended deposit-storage-access.
D. Deposit Compliance	Data content and supporting metadata deposited are checked at the point of deposit for compliance with defined criteria, e.g., data formats, metadata elements, and compliance with legal and ethical norms.
C. Initial Curation	In addition to Level D above, if these criteria are not met the digital objects are curated by the repository to meet the defined criteria.
B. Logical-Technical Preservation	In addition to D and/or C above, the repository takes long-term responsibility for ensuring that the data and metadata can be rendered as required by the designated community.
A. Conceptual Preservation for Understanding and Reuse	In addition to B above, the repository takes long-term responsibility that the data content and metadata can be independently understood by the designated community.

Analysis of CoreTrustSeal Applications 2018–2022

In an effort to better understand the capacity of the proposed model of curation and preservation levels to capture the respective practices of repositories, an analysis of successful CoreTrustSeal applications was carried out. For this purpose, the published application PDFs from applicants awarded the CoreTrustSeal in the years 2018 to 2022 were analysed. The analysis focused on

- the level(s) of curation offered by the applicant as stated under ‘R0. Level of Curation Performed’;
- any additional explanations concerning the level of curation as stated under ‘R0. Level of Curation Performed Comments’;
- the response and compliance level of ‘R8. Appraisal’²;
- the response and compliance level of ‘R10. Preservation Plan’.

The response texts were analysed to determine, among other things,

- the presence/absence of ingest checks against a defined set of deposit criteria (based on R8. Appraisal);

² In this paper, specific CoreTrustSeal requirements will be referred to using their number (e.g., ‘R1.’) and their short name as included in the Requirements publication (e.g., ‘R1. Mission/Scope’).

- the extent of initial curation carried out by the applicant at ingest of resources into the repository;
- the presence of a preservation policy, linked under 'R10. Preservation Plan';
- the preservation strategy employed (format migration, emulation, other/unspecified);
- the presence/absence of different levels of preservation offered by the applicant.

In addition, the certification date and version of CoreTrustSeal Requirements against which the applicant was certified were recorded.

The content analysis and coding were carried out solely based on the responses in the application PDFs. That is, links to evidence were not followed unless this was absolutely necessary to understand the response text. In consequence, the results of the content coding only allow us to draw conclusions about the information provided directly in the responses. For example, the absence of information on format migration or ingest checks in a response does not mean that the applicant does not actually perform these measures — they may simply not have included this information in their response. Thus, the responses have only a limited capacity for enabling us to identify gaps. They are a valuable tool, however, in determining practices and measures that are performed by the repositories.

Description of the Sample

149 application PDFs by 134 applicants were analysed. Fifteen application processes successfully completed in 2021 and 2022 were recertifications. The applicants were previously certified under the 2017–2019 CoreTrustSeal Requirements (Edmunds et al., 2016) and recertified under the 2020–2022 Requirements (CoreTrustSeal Standards and Certification Board, 2019). Figure 1 shows the distribution of applications over the years, differentiated by the version of the CoreTrustSeal Requirements against which the applicant was certified.

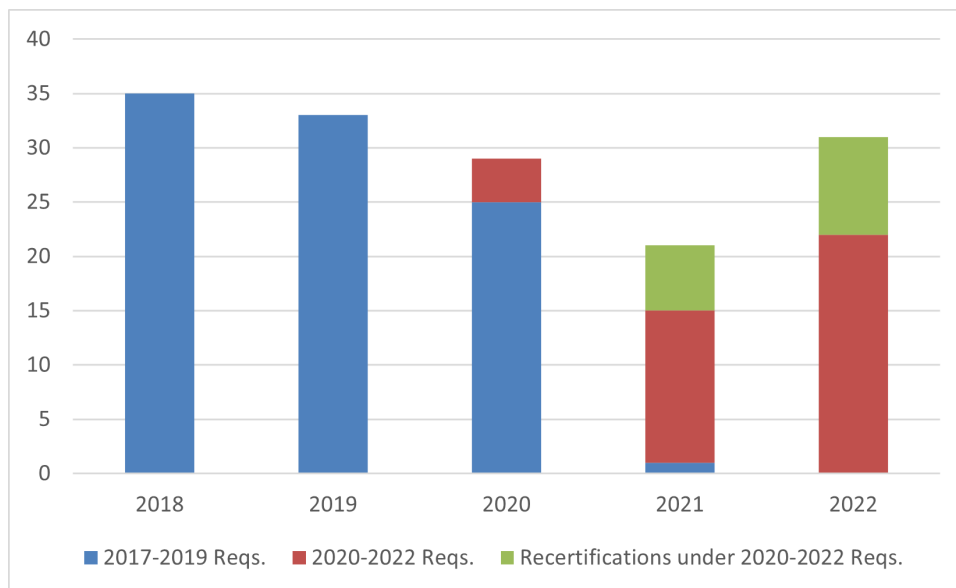


Figure 1. Number of certifications per year, differentiated by version of CoreTrustSeal Requirements (n = 149).

Figure 2 shows the distribution of levels of curation stated in the applications (multiple levels per applicant can be stated), as well as the distribution of the highest levels performed. While

levels of curation B and C were performed most frequently by applicants—mentioned by 65% and 70% of applications respectively—50% of applications named D as the highest level of curation performed. Analysis showed no significant change in this distribution between the 2017–2019 and 2020–2022 Requirements. Focusing on the repositories recertified under the 2020–2022 Requirements, four out of the 15 increased their highest level of curation between certifications.

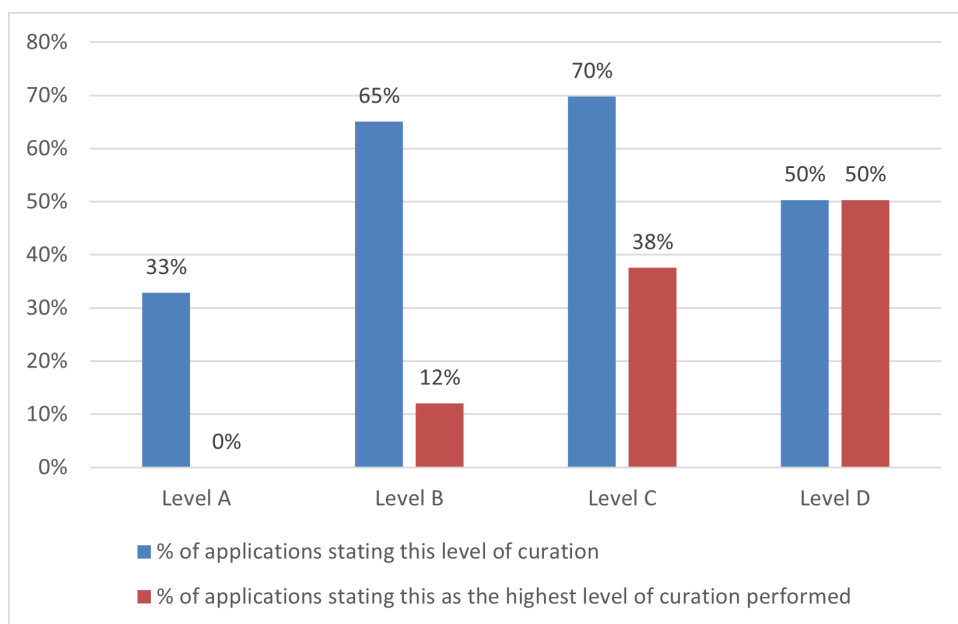


Figure 2. Levels of curation and highest level performed (n=149).

Content Analysis I: Ingest Checks and Initial Curation

The comments applicants provided for ‘R0. Level of Curation Performed’ and the responses to ‘R8. Appraisal’ as well as ‘R10. Preservation Plan’ were analysed to determine which curation measures applicants performed at ingest. All but two applicants provided information in their responses pertaining to human-mediated and/or automatic checks at ingest.

Responses were coded to document whether applicants describe minimal or extensive initial curation measures in their responses. We defined measures as ‘minimal’ if they entailed correction of small errors in the metadata or data by repository staff and providing support for data depositors in carrying out corrections of data and metadata before resubmission, often in the sense of doing these things ‘together’ with the depositors.

Measures were considered ‘extensive’ when they entailed the creation of new metadata or documentation by repository staff, or when data harmonisation and/or format conversions (‘normalisation’) were performed during ingest.

128 of the applications mention measures of initial curation in the analysed responses, thus indicating that they perform curation at ingest beyond merely checking for compliance with deposit criteria and returning the application back to the depositor in case the criteria are not met. As Figure 3 illustrates, not unexpectedly, ‘extensive’ initial curation measures are most frequently offered by applicants who also stated ‘D’ as their highest level of curation offered. However, the illustration also shows that over thirty applications describe ‘minimal’ curation measures at ingest despite selecting curation levels C (17 applications) or D (14 applications) as their highest level of curation, each of which is as per the CoreTrustSeal Requirements associated with a more comprehensive curation of the deposited (meta)data, including normalisation and data corrections.

A possible reason for this may be that the applicants in question accept data that is very uniform and homogeneous, e.g., when directly submitted in an automated process from an

instrument, data that is not particularly complex in terms of structure and format, that does not carry any potential for containing sensitive information, and/or that was created by or involving repository staff to the effect that the data is curated extensively in the pre-repository phase.

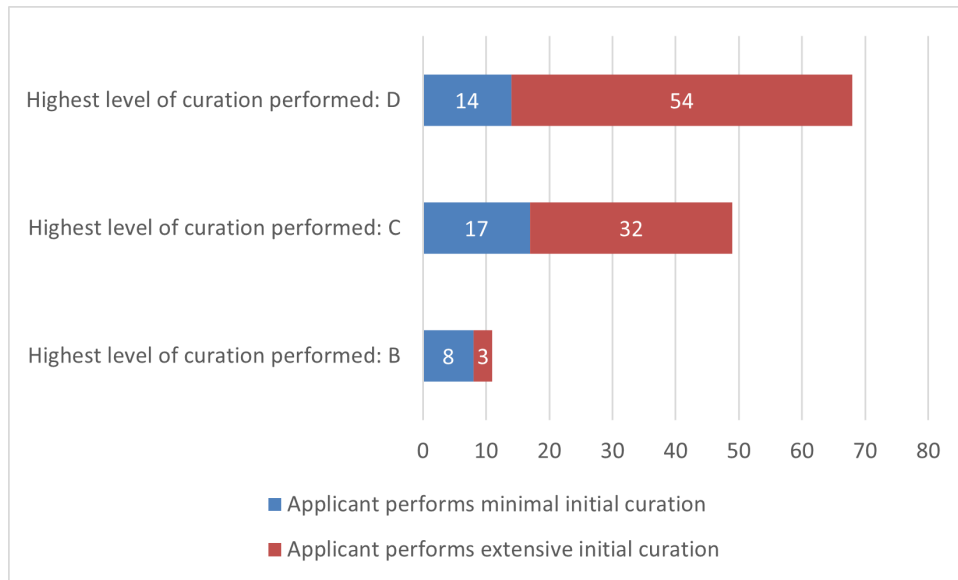


Figure 3. Highest level of curation and extent of initial curation (n=128).

Content Analysis II: Active Preservation and Preservation Levels

Of all analysed applications, 97 (65%) provided a link to a preservation plan as evidence in their responses to 'R10. Preservation Plan'. Of the applications not providing a link, 30 were at Compliance Level 3 (i.e., 'in the implementation phase' for this requirement), leaving 22 applications that did not provide a link despite being assessed at Compliance Level 4 (i.e., 'fully implemented in the repository'). Figure 4 shows that the percentage of applications providing a link to a plan has climbed steadily over the years, suggesting that review practice became stricter and more consistent in this regard.

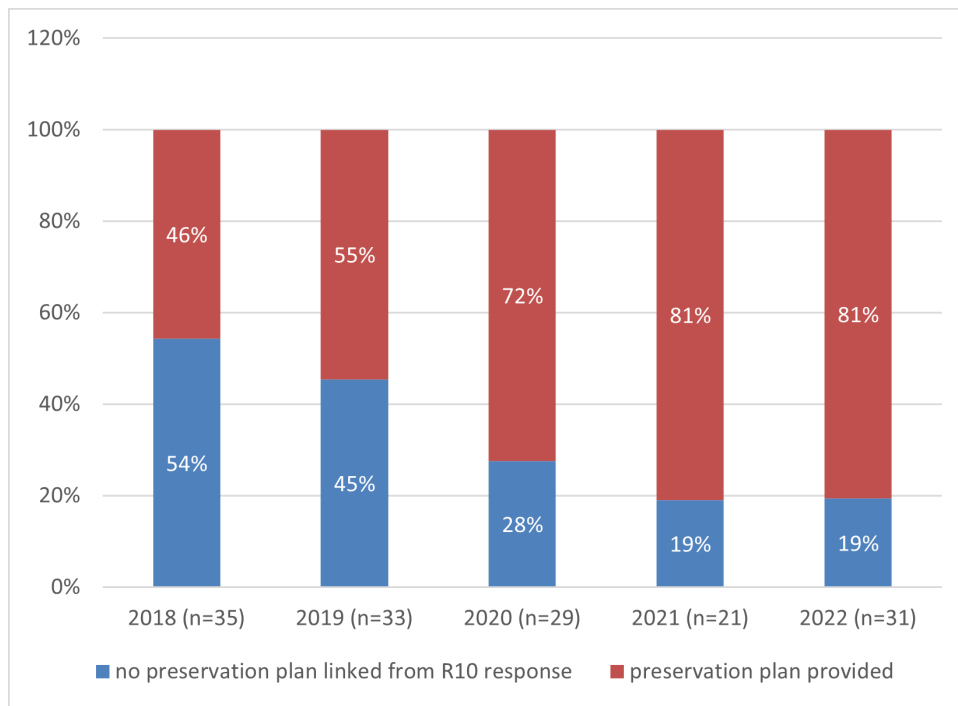


Figure 4. % of applications with linked preservation plans per year of certification.

The responses to ‘R10. Preservation Plan’ were also analysed for mentions of preservation strategies currently employed by the applicant. The responses were coded for references to bitstream preservation, format migration, emulation, and other/unspecified strategies. As shown in Figure 5, 28% of responses (42 applications) did not refer to any preservation strategy. 20% (30 applications) of the remaining responses only described bitstream preservation measures. Thus, 48% of R10 responses make no reference to active digital preservation measures. The preservation strategy mentioned most frequently, format migration, is referred to by all but two R10 responses from the remaining 77 applications. As shown in Figure 6, the percentage of applications mentioning active preservation measures under R10 each year of certification has increased over time. Again, this might indicate that reviewers became stricter and/or applicants made sure to mention the employed preservation strategies in their responses.

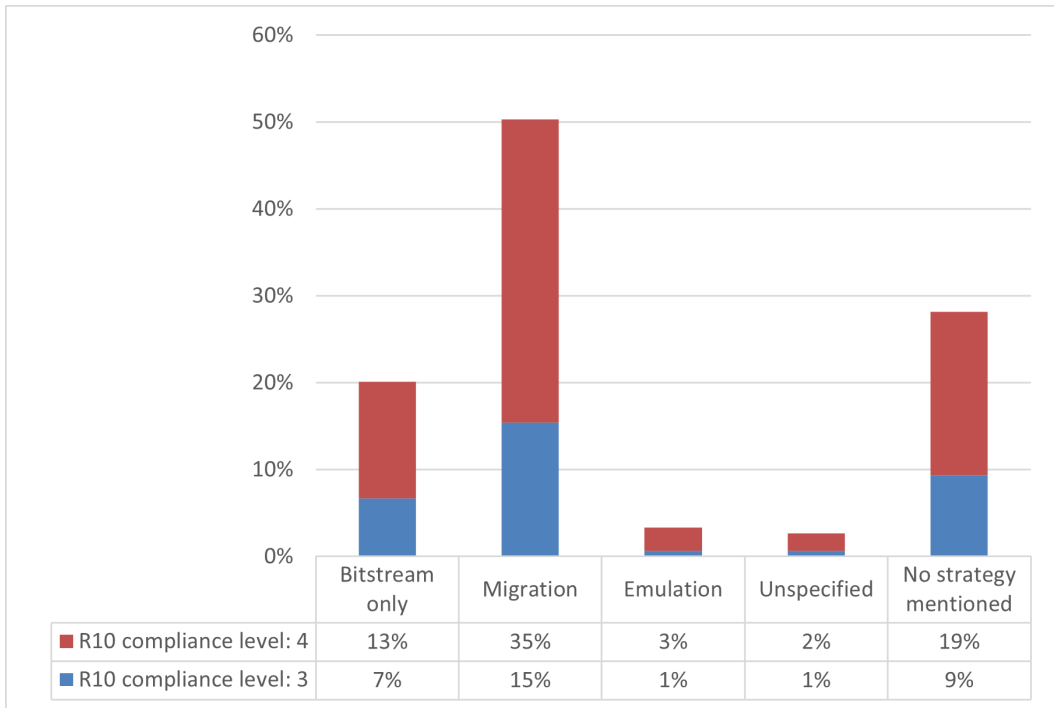


Figure 5. Preservation strategies mentioned in R10 by % of applications (n=149).

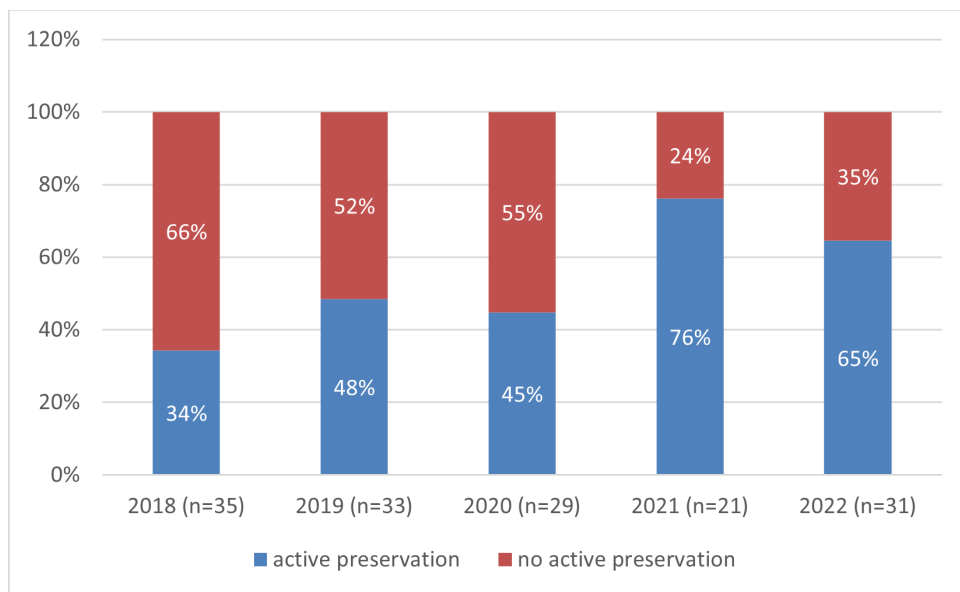


Figure 6. % of applications stating active preservation measures in R10 per year of certification.

The responses to ‘R0. Level of Curation Performed Comments’ and ‘R10. Preservation Plan’ were also analysed for whether applicants described or referenced employing different levels of preservation to different parts of their collections. 21 applicants (recertifications were counted only once as practice did not change between certifications) described employing two or more distinctive levels of preservation (see Figure 7). Out of these, 12 applicants (57%) employed two levels, in most cases identifying bitstream preservation as the base level and an active preservation strategy as second level. Seven applicants reported distinguishing three or more

levels (33%). In these cases, often additional curation measures (such as file format normalisation at ingest, creation of documentation, creation of derived datasets, etc.) were included in the levels, with bitstream preservation and/or functional preservation measures such as format migration forming the base level.

In two cases, the evidence describing the levels of preservation in more detail was no longer accessible, so that neither the number nor type of levels employed could be determined.³

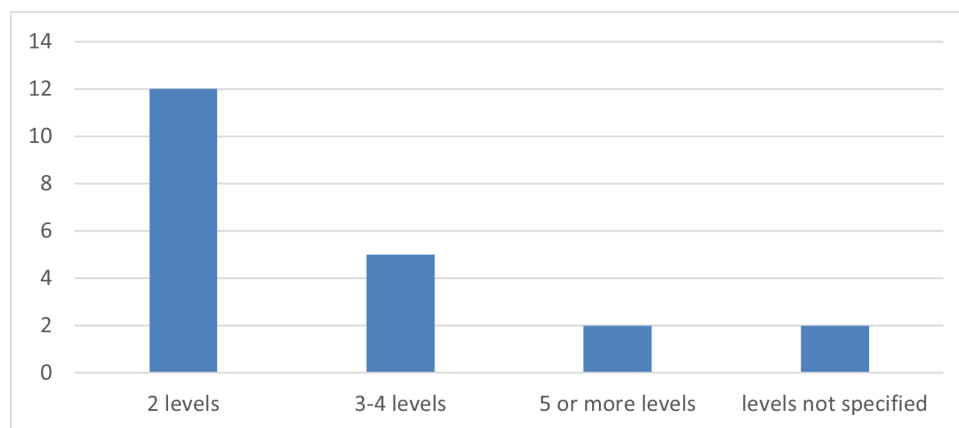


Figure 7. Number of preservation levels by applicant (n=21).

Mapping to Proposed Model of Curation and Preservation Levels

Based on the analysis of responses to ‘R0. Level of Curation Performed’, ‘R8. Appraisal’, and ‘R10. Preservation Plan’, a partial mapping to the proposed model of curation and preservation levels was carried out. ‘Level A. Conceptual Preservation’ from the proposed curation and preservation levels model is not yet included in the mapping. Future analyses of related responses will be undertaken.

For each application, the level of curation activity at ingest and the preservation level was mapped as shown in Appendix A. Figure 8 shows the maximum (combined) levels achieved by applicants. 71 applications only provided information about ingest checks or initial curation without indicating they also perform logical-technical preservation. 77 applications indicated providing ingest checks or initial curation in combination with logical-technical preservation. Based on the analysis and mapping, one application provided only information about an unattended storage-access service without any initial curation or preservation of digital objects (corresponding to the curation level ‘A’ in the CoreTrustSeal Requirements since 2017).

³ In fact, a peripheral finding of our analysis was the wide-spread lack of persistence of supporting evidence for trustworthiness linked from successful CoreTrustSeal applications. Though not unexpected, given that such evidence often resides on organisational websites undergoing frequent changes, the inability to access this evidence now negatively impacts not only our ability to analyse and assess ‘historical’ curation and preservation practices; this lack of application ‘provenance information’ may also affect the perceived trustworthiness / the trust in both the certification process and the certified repositories. Therefore, we recommend that the CoreTrustSeal Board encourages applicants to make any public evidence linked from the application persistently available at the point of certification.

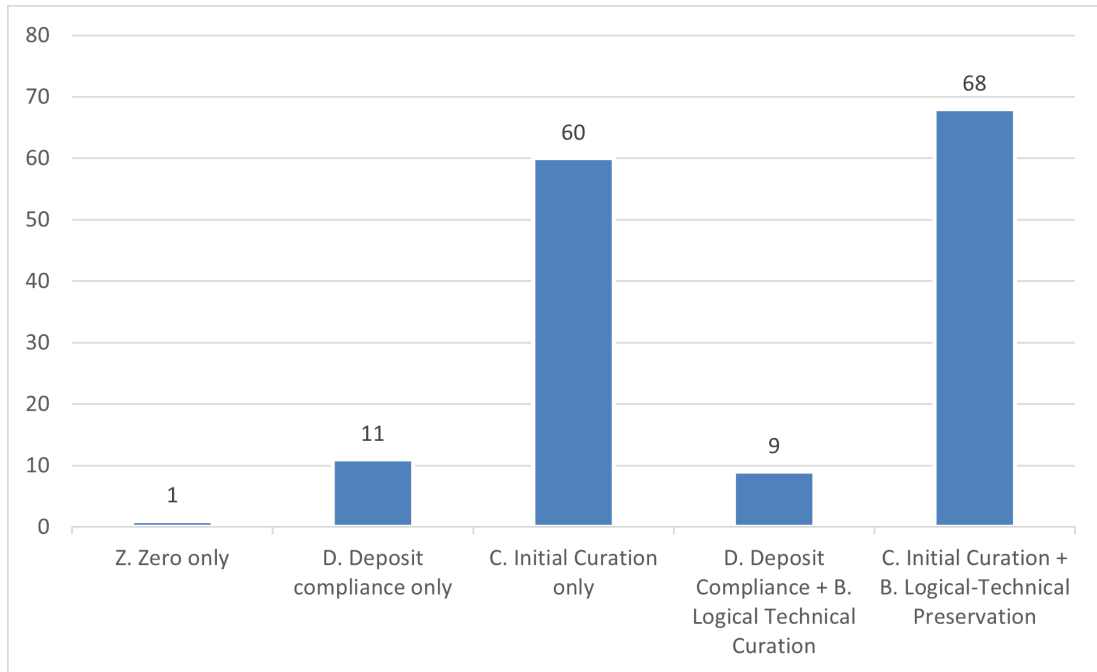


Figure 8. Maximum levels of curation and preservation achieved by applicants (n=149), derived by mapping to new model (see Appendix A).

Conclusion

The current CoreTrustSeal R0 tiers of ‘Level of Curation Performed’ and associated comments provide some context and insight to guide the assessment process. But our analysis indicates that these alone may not be sufficient to define whether an applicant is in scope, by providing active long-term preservation. The current R0 information provides a useful reference point for analysis alongside the other variables we have selected. But overall, we conclude that the proposed model of curation and preservation levels provides clearer and more inclusive information about repository activities. Mapping current catalogues and lists of curation activities such as the *DCN CURATE(D) Steps and Checklist* (Data Curation Network, 2022) and the *ARL SPEC Kit 354* (Hudson-Vitale et al., 2017) to these proposed levels would provide further support and guidance for repositories seeking to implement curation and preservation services. This level of additional mapping would be necessary and beneficial as many of the specific activities taken at the point of deposit through initial curation may be similar or identical to those taken over time to achieve preservation at levels B or A in the proposed CoreTrustSeal Curation and Preservation Levels. The differentiating factor becomes the degree to which a repository undertakes to monitor technologies and communities over time to guide future preservation actions.

The implementation of the proposed levels within the CoreTrustSeal Requirements would reduce ambiguity during self-assessments and facilitate in-scope/out-of-scope decisions about applications. Any decision to place a repository out of scope for certification would be informative about current gaps in assessment and certification services. More broadly, the application of the levels would support structured planning of curation and preservation measures and clarify the levels of care offered by the repository to depositors, users, and funders — especially if paired with similarly structured information about relevant characteristics of the curated/preserved digital objects, and the composition, needs, and knowledge base of the designated community. If adopted, the levels have the potential to be expanded from initial tiers to more explicit, machine-actionable metadata-driven statements about repository services, including retention periods, appraisal criteria (automated and human-mediated), and specific

preservation techniques. The clear corollary to this would be that the digital objects themselves could be supported by metadata that describes the level of care they receive, including any minimum periods of retention or preservation before reappraisal.

In addition to supporting the streamlining of assessment and certification processes, taken together, these measures could facilitate the clear specification of repository responsibilities and the design of curation and preservation plans, thereby increasing the overall transparency of the levels of curation in place across repositories and objects, all the while contributing to the endeavour of better understanding the costs, benefits, and impact of digital preservation in the long term.

References

- Consultative Committee for Space Data Systems. (2012). *Reference Model for an Open Archival Information System (OAIS). Recommended Practice*. (Magenta Book CCSDS 650.0-M-2). Retrieved from <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- CoreTrustSeal Standards and Certification Board. (2019). *CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022*. (v02.00-2020-2022). Zenodo. <https://doi.org/10.5281/ZENODO.3638211>
- CoreTrustSeal Standards and Certification Board. (2022). *CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025*. (V01.00). <https://doi.org/10.5281/zenodo.7051011>
- CoreTrustSeal Standards and Certification Board. (2023). *Curation & Preservation Levels: CoreTrustSeal Discussion Paper* (v02.00). Zenodo. <https://doi.org/10.5281/ZENODO.8083359>
- Data Curation Network. (2022). *CURATE(D) Steps and Checklist for Data Curation* (Version 2). <http://z.umn.edu/curate>
- Edmunds, R., L'Hours, H., Rickards, L., Trilsbeek, P., Vardigan, M., & Mokrane, M. (2016). *Core Trustworthy Data Repositories Requirements*. Zenodo. <https://doi.org/10.5281/ZENODO.168411>
- Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation* 3(1), 134–140. <https://doi.org/10.2218/ijdc.v3i1.48>
- Hudson-Vitale, C., Imker, H., Johnston, L., Carlson, J., Kozlowski, W., Olendorf, R., & Stewart, C. (2017). *SPEC Kit 354: Data Curation* (May 2017). Association of Research Libraries. <https://doi.org/10.29242/spec.354>
- International Organization for Standardization. (2012). *ISO 16363:2012. Space data and information transfer systems—Audit and certification of trustworthy digital repositories* (ISO 16363:2012). <https://www.iso.org/standard/56510.html>
- Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C., Blake, M., Herndon, J., McGeary, T. M., & Hull, E. (2018). Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data. *International Journal of Digital Curation* 13(1), 125–140. <https://doi.org/10.2218/ijdc.v13i1.616>

- Lafferty-Hess, S., Rudder, J., Downey, M., Ivey, S., Darragh, J., & Kati, R. (2020). Conceptualizing Data Curation Activities Within Two Academic Libraries. *Journal of Librarianship and Scholarly Communication* 8(1), Article 1. <https://doi.org/10.7710/2162-3309.2347>
- nestor Certification Working Group. (2013). *Explanatory Notes on the nestor Seal for Trustworthy Digital Archives* (nestor-materials 17; Version 1). <https://nbn-resolving.org/urn:nbn:de:0008-2013100901>
- PREMIS Editorial Committee. (2015). *PREMIS Data Dictionary for Preservation Metadata. Version 3.0*. <https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>
- Recker, J., L'Hours, H., & Kleemola, M. (2023). Modelling curation and preservation levels for trustworthy digital repositories. *Proceedings of PV2023: Adding Value (to) and Preserving Scientific and Technical Data. 2-5 May 2023, CERN*. <https://doi.org/10.5281/zenodo.10052805>
- Research Libraries Group. (2002). *Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report*. Retrieved from <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>
- Welter, D., Juty, N., Rocca-Serra, P., Xu, F., Henderson, D., Gu, W., Strubel, J., Giessmann, R. T., Emam, I., Gadiya, Y., Abbassi-Daloui, T., Alharbi, E., Gray, A. J. G., Courtot, M., Gribbon, P., Ioannidis, V., Reilly, D. S., Lynch, N., Boiten, J.-W., ... Burdett, T. (2023). FAIR in action—A flexible framework to guide FAIRification. *Scientific Data* 10(1), 291. <https://doi.org/10.1038/s41597-023-02167-2>

Appendix A

Source variable and description	Target variable and description
<p>r0_loc_a: The repository carries out curation on Level A. Content distributed as deposited.</p>	<p>new_zero: The applicant's ingest practices conform to the newly proposed Level Z. Zero: Content distributed as deposited. Unattended deposit-storage-access.</p>
<p>ingest_checks: The repository carries out ingest checks to assess deposit compliance. Based on the responses to 'R0. Level of Curation Performed Comments, and R8. Appraisal'.</p>	<p>new_d: The applicant's ingest practices conform to the newly proposed Level D. Deposit Compliance: Data content and supporting metadata deposited are checked at the point of deposit for compliance with defined criteria.</p>
<p>min_initial_curation: The applicant provides minimal initial curation, e.g., correction of small errors in the metadata or data; providing advice / support to the data depositors in carrying out corrections of data and metadata before resubmission. Based on the responses to 'R0. Level of Curation Performed Comments', and 'R8. Appraisal'.</p>	<p>new_c: The applicant's ingest practices conform to the newly proposed Level C. Initial Curation: In addition to Level D, if these criteria are not met, the digital objects are curated by the repository to meet the defined criteria.</p>
<p>extensive_initial_curation: The applicant provides extensive measures of initial curation, e.g., creation of metadata and documentation, data harmonisation, format conversions. Based on the responses to 'R0. Level of Curation Performed Comments', and 'R8. Appraisal'.</p>	<p>new_c: see above</p>
<p>format_migration: The response to 'R10. Preservation Plan' indicates that the applicant performs format migrations to mitigate the threat of file format obsolescence.</p>	<p>new_b: The applicant's ingest practices conform to the newly proposed Level B. Logical-Technical Preservation: In addition to D and/or C above, the repository takes long-term responsibility for ensuring that the data and metadata can be rendered as required by the designated community.</p>
<p>emulation: The 'R10. Preservation Plan' indicates that the applicant employs emulation to mitigate the threat of technical obsolescence.</p>	<p>new_b: see above</p>
<p>other_strategy: The 'R10. Preservation Plan' mentions a different or unspecified digital preservation strategy.</p>	<p>new_b: see above</p>