

## **Administrative Metadata for Long-Term Preservation and Management of Resources: A Survey of Current Practices in ARL Libraries**

Rutgers University has made this article freely available. Please share how this access benefits you.  
Your story matters. [\[https://rucore.libraries.rutgers.edu/rutgers-lib/41069/story/\]](https://rucore.libraries.rutgers.edu/rutgers-lib/41069/story/)

This work is an **ACCEPTED MANUSCRIPT (AM)**

This is the author's manuscript for a work that has been accepted for publication. Changes resulting from the publishing process, such as copyediting, final layout, and pagination, may not be reflected in this document. The publisher takes permanent responsibility for the work. Content and layout follow publisher's submission requirements.

Citation for this version and the definitive version are shown below.

**Citation to Publisher** Otto, Jane. (2014). Administrative Metadata for Long-Term Preservation and Management of Resources: A Survey of Current Practices in ARL Libraries. *Library Resources and Technical Services* 58(1).  
**Version:**

**Citation to *this* Version:** Otto, Jane. (2014). Administrative Metadata for Long-Term Preservation and Management of Resources: A Survey of Current Practices in ARL Libraries. *Library Resources and Technical Services* 58(1). Retrieved from [doi:10.7282/T3R49NRF](https://doi.org/10.7282/T3R49NRF).

**Terms of Use:** Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

*Article begins on next page*

### Administrative Metadata for Long-Term Preservation and Management of Resources: A Survey of Current Practices in ARL Libraries

PLEASE NOTE: This work is an **ACCEPTED MANUSCRIPT (AM)**

This is the author's manuscript for a work that has been accepted for publication. Changes resulting from the publishing process, such as copyediting, final layout, and pagination, may not be reflected in this document. The publisher takes permanent responsibility for the work. Content and layout follow publisher's submission requirements.

Citation for this version and the definitive version are shown below.

<b>Publisher Version</b>	"Administrative Metadata for Long-Term Preservation and Management of Resources: A Survey of Current Practices in ARL Libraries," <i>Library Resources and Technical Services</i> 58: 1 (January 2014).
<b>Terms of Use</b>	Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

**Rutgers Community Repository (RUcore)** is a repository of digital research and educational materials created and used by the University community and its strategic collaborators. The goal of the Rutgers University Community Repository is to advance research and learning at Rutgers, to foster interdisciplinary collaboration, and to contribute to the development of new knowledge through the archiving, preservation, and presentation of digital resources.

↻ Article begins on next page ↻

# **Administrative Metadata for Long-Term Preservation and Management of Resources: A Survey of Current Practices in ARL Libraries**

By Jane Johnson Otto

Jane Johnson Otto (jjotto@rulmail.rutgers.edu) is Media and Music Metadata Librarian, Technical and Automated Services, Rutgers University Libraries.

## **Abstract**

*An institutional repository is, among other things, a means to preserve an organization's scholarly output or resources in a variety of digital media and across disciplines. Administrative metadata is critical to the preservation of these digital resources. This study, which surveyed fifty-four Association of Research Libraries (ARL) institutional repositories about their administrative metadata, was designed to create a snapshot of current metadata practices. It revealed no true consensus of administrative metadata accommodated and collected by the repositories. Moreover, responses throughout the survey indicate that in general, organizations are neither accommodating nor recording administrative metadata to any significant extent. If research libraries are to provide permanent, organized and secure repositories for institutional scholarship and special collections, they must identify core metadata in the context of repository objectives, explore barriers to collection of administrative metadata, and strategize as to how those barriers might be mitigated or overcome.*

## **Introduction**

An institutional repository is a central digital repository for an organization's scholarly output across media and disciplines. It is organized and secure, and the digital objects (scholarly resources) it houses are intended to be permanently preserved. A scan of Association of Research Libraries (ARL) repository websites shows that many libraries make explicit this "preservation promise" to depositors and other users.

Administrative metadata, which describes the technical characteristics of the digital file and any original physical source object, preservation actions, and relevant intellectual property rights and access permissions, is critical to the preservation of digital resources.<sup>1</sup> Ten years ago, early in the development of institutional repositories, lack of preservation and administrative metadata was cited as the biggest obstacle to successful long-term preservation.<sup>2</sup> Six years ago, the *Audit & Certification Criteria and Checklist* developed by the Research Libraries Group-National Archives and Records Administration (RLG-NARA) Task Force on Digital Repository Certification addressed metadata again, and suggested “preservation metadata is best addressed by members of the designated communities.”<sup>3</sup>

These communities have responded to the call by developing numerous standards to support the preservation and management of digital objects. A review of these standards reveals that administrative metadata is detailed and voluminous, and for good reason. Most librarians and archivists would agree that the more information that is known about a resource, the more effectively it can be managed and preserved. Unfortunately, the gathering, recording, and management of detailed metadata is expensive; indeed, a 2002 report by RLG stated that creation of detailed *technical* metadata alone was possibly beyond the human resources of most institutions.<sup>4</sup> Ten years later, this is still the case.

To further complicate matters, the array of standards, best practices, and models fail to form a cohesive whole. As if the sheer volume of detailed administrative metadata were not daunting enough, the overlap and gaps between the related standards make them difficult to implement within any repository system. There are preservation metadata standards meant to apply to all formats, but which lack the specificity needed for any one.<sup>5</sup> There are technical metadata standards intended for a specific format but which lack corresponding metadata for the original (usually analog) source material.<sup>6</sup> Those technical metadata standards which do include source metadata are often only applicable to one format.<sup>7</sup> Moreover, boundaries between metadata types are not clearly delineated. Some preservation standards include rights metadata,<sup>8</sup> but some do not. Finally, there are standards that were not designed to record detailed administrative metadata, but which have nonetheless been extensively used for this purpose.<sup>9</sup>

Within this complex landscape lives a community of institutions often overwhelmed by an influx of digital resources and struggling to balance the value added by metadata and the significant cost of that metadata's creation and maintenance. Individual institutions may have a good idea of how much metadata it is practical to record based on their individual circumstances; what the community needs is a clearer picture of what administrative metadata is critical to the preservation mission, particularly in the increasingly collaborative information space. As Chen and Reilly stated, "one of the biggest challenges in preservation automation is to develop a strategic preservation metadata plan *and decide how much information we need to record* and whether the information can be accurately recorded."<sup>10</sup> [emphasis added by author] At this point it is unclear what metadata is actually being collected for repositories, or even what metadata these established repositories are able to support.

This article provides the first snapshot of current practices through a survey that identifies administrative metadata *accommodated* by ARL repositories and administrative metadata actually *collected* for those repositories. This study assumes that digital preservation requires administrative metadata and was designed to answer these questions:

1. What administrative metadata *can* repositories collect (given their current schemas?)
2. How much administrative metadata is actually *being* collected?
3. What are the common elements of administrative metadata collected across a majority of ARL institutional repositories?
4. Is the metadata perceived to be sufficient to support the tasks a repository is expected to perform?

The survey was intended to determine the extent to which repositories are able to accommodate robust administrative metadata, and the degree to which organizations are collecting it. Responses should help identify any gaps between metadata currently (and/or commonly) collected, and metadata needed to support digital preservation. It is hoped that survey results will inform discussions of the current state of digital preservation, and ways to move forward.

## Literature Review

Digital preservation has been identified as a primary attribute and responsibility of the trusted repository in numerous seminal publications, notably the 2002 RLG-OCLC report on trusted digital repositories,<sup>11</sup> Clifford Lynch's 2003 article on institutional repositories (IRs),<sup>12</sup> the 2003 Joint Information Systems Committee (JISC) report on e-prints preservation,<sup>13</sup> the 2006 Nestor Working Group *Catalogue of Criteria for Trusted Digital Repositories*<sup>14</sup>, and the 2007 RLG-NARA *Audit & Certification Criteria and Checklist*.<sup>15</sup> Several studies suggest digital preservation is valued by libraries, repository contributors, and users.<sup>16</sup>

These publications notwithstanding, the professional literature still suggests that repositories and librarians are unprepared to preserve their digital objects and, for the most part, have not yet done so. Ten years ago, the 2003 *Invest to Save* report enumerated the "severe limitations" to digital preservation methods, processes, strategies, systems and technologies, warning of "great risk that valuable digital content will not survive for the long term."<sup>17</sup> The 2003 JISC report on e-prints preservation described the field in a state of flux and uncertainty; repository managers had yet to fully engage with preservation challenges, and were "unsure of how to proceed"; this uncertainty was echoed several years later in responses to the 2006 *Census of Institutional Repositories in the United States*.<sup>18</sup> Ross in 2003 found that few organizations were actively developing digital preservation solutions, preservation models tended to be reactive and ad hoc, and few organizations seemed aware of the complexities with migration or the enormity of the preservation problem.<sup>19</sup> In 2005 Knight listed a number of "significant concerns as to how a sustainable outcome will be achieved in this arena," including low-level awareness of need, as well as a lack of metrics (regarding the scope of the challenge), skill sets, agreed approaches, practical models, and collaboration.<sup>20</sup> A 2006 Canadian Association of Research Libraries (CARL) institutional repository assessment questioned whether surveyed repositories possessed sufficient resources and expertise to follow through on their posted preservation policies;<sup>21</sup> a study of repository websites that same year found virtually no evidence of long-term preservation plans.<sup>22</sup> DigitalPreservationEurope's (DPE) 2007 Research Roadmap pointed to a lack of systematic approach to preservation, "no common understanding

of the precise definition of digital preservation” and a “lack of adequate knowledge transfer.”<sup>23</sup> An additional factor is cost. Several authors have noted the difficulty of predicting preservation costs<sup>24</sup>, but from the outset it has been clear that digital preservation requires a steady long-term commitment of resources.<sup>25</sup>

Most of these issues apply equally to the creation and maintenance of the administrative metadata that supports the digital preservation process. Although metadata is just one component of this process, its role is a critical one. The 2003 JISC report called preservation metadata the blueprint for the preservation strategy, and provided ample justification for all types of administrative metadata.<sup>26</sup> Knight cited a “genuine uncertainty as to when preservation metadata is to be captured, how it will be captured, who updates it, and when.”<sup>27</sup> Alemneh’s research indicates that the most frequently identified barriers to the adoption of PREservation Metadata: Implementation Strategies (PREMIS) include lack of training/expertise and perceived lack of knowledge necessary to be confident in the ability to implement PREMIS.<sup>28</sup> Caplan has noted that preservation metadata is not simple to understand, obtain or implement, and characterizes preservation metadata as “a repository’s best guess” as to the information needed to enable use of a resource into the future.<sup>29</sup> Dappert and Enders describe the complexities associated with standard schemas, and the difficulty of striking “the right balance between generality and specificity.”<sup>30</sup> Dappert and Farquahar (2009) assert that current metadata dictionaries are still vague and “await increased practical experience to establish the proper level of granularity.”<sup>31</sup> And again, there is the issue of cost. Metadata has been called “one of the most costly aspects of digital preservation.”<sup>32</sup> Metadata extraction tools show promise, but only limited categories of preservation metadata can be extracted, and comparative analyses of the tools have revealed some shortcomings.<sup>33</sup>

Given these barriers to effective metadata creation, and particularly given the concerns of metadata and digital preservation costs, it is surprising that there is virtually no data about what, or even how much, administrative metadata is actually being collected, to determine how that metadata supports the preservation aspect of the repository mission. Most surveys on repository metadata concern descriptive metadata, or discuss metadata generally, without delving into the specifics of administrative

metadata.<sup>34</sup> Li and Banach surveyed ARL libraries about digital preservation of institutional repository materials, but asked broader questions of policies, strategies, rights, content quality, and sustainability. In ARL's own spec kit on digital preservation, metadata questions were general and few.<sup>35</sup>

A better picture of actual repository capabilities and practices should yield more concrete answers to questions of digital preservation cost and the ability of current practices to support the preservation mission. Related questions about staffing, preservation tasks supported by the metadata, and perceptions of the metadata's adequacy, could shed light on other issues raised in the literature. Finally, it was hoped that the survey might reveal commonalities that could stimulate discussion about guidelines and collaborative uses for administrative metadata, to promote the longevity of digital collections.

## **Methodology**

Assessing the value of any particular type of metadata is impossible without knowing the mission of the repository. Therefore, considerable preliminary research was conducted on two fronts: first, to identify and define discrete categories of metadata and individual elements belonging to each metadata type, and second, to identify and distill into a single list the multitude of preservation-related tasks considered important for digital repositories, as established in the literature. This knowledge guided formulation of the survey questions as well as the response options. The survey was drafted using SurveyMonkey; an Institutional Review Board (IRB) waiver was obtained, and the survey was tested and timed by a number of colleagues nationwide.

ARL is a nonprofit organization of 126 libraries at comprehensive, research-extensive institutions in the United States and Canada, which share similar missions, aspirations, and achievements. Participation was limited to these libraries in order to provide a focused and representative sample of research organizations which host digital repositories and that generally work collaboratively to preserve and make available the scholarly record.

Repository contacts comprising the survey sample were drawn from the Registry of Open Access Repositories (ROAR) and the Directory of Open Access Repositories (OpenDoar)<sup>36</sup>, as well as from the



institutions' websites. Participants were asked to provide a single institutional response. If an organization had more than one repository, the participant was asked for a response pertaining to "your primary repository, the one housing your organization's scholarship, and/or digital collections, i.e., the one most aligned with your institutional mission." Those organizations hosting a consortial repository were asked to coordinate a single response from a representative of that repository.

In May, 2012, 104 survey invitations were sent and fifty-five complete responses were received. Of these, one response represented a repository still under development, for which metadata had not yet been finalized, and that survey response was excluded from the resulting data set, for a completion rate of 52 percent of the sample.

## **The Survey**

The confidential survey consisted of thirty-seven questions about the repository, its administrative metadata schema, individual metadata elements, and repository tasks that metadata supports. "Metadata schema" is used to refer to an organized and documented set of metadata elements. A schema may be internal to an organization or may represent a shared metadata standard, which is managed by a standards body and open to community review and reuse.

Because of the diversity of metadata standards and definitions, the scope of the questions were limited to four types of administrative metadata that directly impact preservation, defined as follows:

- *Rights metadata.* Information about intellectual property rights granted or reserved, copyright holder or licensor, etc.
- *Technical metadata.* Metadata describing the characteristics of the archival digital file, e.g., file size, compression scheme, operating system, codec, etc.
- *Preservation metadata.* Metadata supporting the digital preservation process, beyond that found in technical metadata

- *Source metadata.* Metadata documenting the physical characteristics of the original (usually analog) *physical* source object from which the digital master is derived (for example, an original film negative or vinyl record), e.g., dimensions, sound and color characteristics, etc.

Respondents were asked a series of questions for each metadata type, in turn: first, rights metadata, then technical, next preservation, and last, source. For each type, it was first determined if the respondent's repository accommodated that type. If the answer was "yes," the respondent was asked several more questions pertaining to that type of metadata. If the answer was "no," the respondent skipped forward to a similar series of questions for the next metadata type. The entire survey is appended to this paper.

## **Survey Results**

### **The repositories**

Although repository names cannot be published due to IRB restrictions and agreements made with respondents, it can be said that 83 percent of the fifty-four respondents answered the survey *either* in terms of "*the institutional repository*" ("a repository of broad scope but limited to the organization's scholarly output") (41 percent) *or* in terms of a repository "combining scholarly output with digitized library archive collections" (43 percent). Thirteen percent of respondents answered questions in reference to "the organization's digital library, limited to digitized library or archive collections."<sup>37</sup> One respondent spoke on behalf of a consortial repository; another described her repository as a combination of digitized graduate theses and scholarly output of library [i.e., not university] faculty. Repositories were equally divided between small (under 1,000 fully cataloged resources added annually), medium (1,000-5,000 added annually), and large (over 5,000). The most heavily used repository software was DSpace, employed by 30 respondents (56 percent), followed by Fedora and custom/in-house applications (six institutions, or 11 percent, each).

Most of the surveyed organizations accommodated textual materials, still images, video, and audio. Nearly three-quarters accommodated data sets. Those checking the "other" box listed websites

(four respondents), code or software (three), musical compositions (one), and “Encoded Archival Description (EAD) finding aids, MARC21 collection-level records, and PDF inventories linked to collection-level records” (1).<sup>38</sup>

### **Administrative metadata employed**

Respondents were presented with a list of fifteen metadata standards and asked to check all from which they had incorporated any elements for rights, technical, source, or preservation metadata; additional standards could be specified under “other.”<sup>39</sup>

Twelve of these fifteen metadata standards, plus an additional twelve specified under “other” (for a total of twenty-four), were incorporated into the metadata schema of at least one repository.<sup>40</sup> The majority of respondents (59 percent) employed more than one metadata standard.<sup>41</sup> On average, repositories combined two. One organization combined eleven standards, but twenty-two organizations (42 percent) used just one.

The three most heavily used metadata standards were Qualified Dublin Core (81 percent), Simple Dublin Core (43 percent), and Metadata Object Description Schema (MODS) (30 percent). Moreover, each of the twenty-two libraries employing just one standard used Dublin Core (most frequently, Qualified Dublin Core) or MODS.<sup>42</sup> Ironically, each of these three standards was originally designed for descriptive metadata.

The range of metadata standards incorporated into any one repository’s schema, and the spread across standards, suggests a lack of consensus regarding the viability of any one standard or series of standards.

[insert table 1]

## Rights metadata

Over three-quarters of repositories (forty-two, or 78 percent) accommodate some rights metadata. Of the twelve respondents whose repositories do not accommodate rights metadata, two indicated rights for their resources are known and fall into a single (or small number of) rights categories, suggesting the information could be kept in institutional memory. Another noted it tracks licenses for repository resources, but not via repository metadata.<sup>43</sup>

Since in all probability, different metadata is recorded for different types of objects and formats (faculty scholarship, research data, video, etc.), it was not practical to ask respondents exactly what metadata is routinely recorded across all types of repository objects. Therefore, the survey asked which metadata elements are *accommodated by the repository schema*, and as a corollary, for what portion of the repository objects (roughly) is *some* of that metadata actually recorded.

Respondents were presented with a list of eighteen rights elements and asked to check those accommodated by their repository metadata schema. (It is important to keep in mind throughout the survey that respondents were directed to select an element from the list only if their metadata schema “has an element dedicated to that information, or a more granular form of it.” When a respondent left a metadata element unchecked, it does not mean that metadata cannot be recorded in the repository; it simply means there is no element *dedicated* to that metadata. In other words, the metadata is not sufficiently parsed to allow efficient retrieval, machine processing, reporting, and sharing.)

Every element offered on the list was used by at least two repositories. The average number of rights elements accommodated was 4.6. The four most common are:

- 1) Rights statement or license terms (general) (90 percent)
- 2) Copyright status (e.g., copyright protected, public domain) (57 percent)
- 3) Rights granted the repository (replicate, migrate, modify, use, delete, etc.) (48 percent)
- 4) Availability status (e.g., open, restricted, unavailable) (36 percent)

[insert figure 1]

Respondents were then asked to specify the portion of the repository's objects for which rights metadata is actually recorded. Of the forty respondents whose repositories accommodate some rights metadata and who were able to answer this question (two respondents responded "I don't know"), sixteen (40 percent) said they record some of this rights metadata for all of their repository objects. Eight respondents (20 percent) record some of this rights metadata for less than one third of their repository objects, and one of those eight respondents records no rights metadata at all. When all respondents are taken into account, nearly a quarter of repositories (thirteen, or 25 percent) record no rights metadata.<sup>44</sup> Clearly, only a relatively small percentage of repositories are routinely recording sufficient rights metadata.

[insert figure 2]

Next, respondents were asked to gauge their satisfaction with the amount of rights metadata that can be collected (i.e., the number of elements accommodated by the repository) and the amount that actually is collected. Each response on the scale was assigned a value from one (way too little) to five (way too much), with intermediate values of two, three (just right), and four. "I don't know" had a rating of zero. ("I don't know" responses were discarded for this question and others where it could skew results.) Rights metadata capability ("rights metadata you CAN collect") had a higher rating average (2.66) than actual rights metadata practice ("rights metadata you DO (routinely) collect") (2.23).

Thirty-eight respondents assessed the amount of rights metadata the repository can accommodate. Of those, 21 (55 percent) felt the amount of metadata accommodated by their schema is "just right." In terms of actual practice, only seventeen respondents (42 percent) felt the amount of metadata they routinely collect is "just right." Well over half the respondents (twenty-three, or 57 percent) felt the metadata they routinely collect falls on the (less than) "just right" end of the scale, with eight (20 percent) saying the amount was "way too little."

[insert figure 3]

Survey responses were also reviewed individually to identify: 1) how many respondents perceived a gap (in terms of adequacy) between metadata accommodated and metadata routinely supplied, and 2) the size of that gap. For 64 percent of respondents, there was no gap. Twenty-five percent of respondents had a gap of one rating point, and 11 percent had a gap of two rating points.

Finally, data were analyzed to determine if the rating for a repository's rights metadata *capability* correlated to the number of metadata elements the schema accommodates (as derived from question 12). Respondents were divided into groups of roughly equal size, based on the number of metadata elements accommodated, and ratings were compared for each category.

[insert table 2]

In general, the more metadata elements a schema accommodates, the more likely respondents are to rate the metadata they *can* collect as 3 ("just right") or 4 (somewhere between "just right" and "way too much"). Those schemas with the most elements (9-12) were rated 3 or higher by 100 percent of respondents, whereas schemas with the least elements (1) were rated 3 or higher by only 43 percent of respondents.

[insert figure 4]

Interestingly, there appears to be no consensus as to how many elements would warrant a "just right" rating. For the twenty-one who rated the repository capability "just right," the number of elements ranged from one at one end of the spectrum to eighteen at the other end. The disparities found in responses to this question, together with the relatively high percentage of "I don't know" responses (11

percent), indicate there is little consensus as to how much is too much rights metadata, and how much is too little.

[insert table 3]

### **Technical metadata**

Forty-three repositories (80 percent) accommodate some technical metadata (roughly equivalent to the percentage of repositories accommodating some rights metadata). Of the eleven respondents whose repositories do not accommodate technical metadata, one noted “so far the vast majority of materials are pdfs” (suggesting the technical metadata could be kept in memory), and another noted that technical information can be added in a separate document and attached.

Respondents were presented with a list of thirteen technical metadata elements and asked to check those accommodated by their repository metadata schema. Every element offered on the list was used by at least one repository. The average number of technical metadata elements accommodated was 3.9.<sup>45</sup> The five most common are:

- 1) Format (.pdf, .htm) (98 percent)
- 2) File size (88 percent)
- 3) Fixity check data (56 percent)
- 4) Creating application name [and/or version] (34 percent)
- 5) Technical metadata notes (27 percent)

[insert figure 5]

Respondents were next presented with a list of fifteen technical metadata elements specific to video and asked to check those accommodated by their repository metadata schema. Of forty-three respondents offered this question, thirty-seven (86 percent) accept video in their repositories, yet twenty-

two of those (51 percent) answered “none of the above” or some equivalent. In fact, taking all surveyed repositories into account, 48 (89 percent) accept video, yet only seventeen of those (35 percent) accommodate video technical metadata. Therefore, responses to this question (or lack thereof) may say more about the dearth of video metadata than about what metadata is considered useful. Each of the fifteen metadata elements listed was nonetheless accommodated by at least one, and as many as thirteen, repositories.

[insert figure 6]

Respondents were then presented with a list of eleven technical metadata elements specific to audio and asked to check those accommodated by their repository metadata schema. Of forty-three respondents offered this question, thirty-eight (88 percent) accept audio in their repositories, yet nearly half of those (eighteen, or 47 percent) answered “none of the above” or some equivalent. Taking all surveyed repositories into account, forty-seven (87 percent) accept audio, yet only eighteen of those (38 percent) accommodate audio technical metadata. Again, responses to this question reveal more about the dearth of audio metadata than about what metadata is considered most useful. Still, each of the eleven metadata elements listed was accommodated by at least two, and as many as fourteen, repositories.

[insert figure 7]

Respondents were asked about the portion of the repository’s objects for which technical metadata is actually recorded. Of the thirty-eight respondents whose repositories accommodate some technical metadata and who were able to answer this question, twenty (53 percent) record some of this technical metadata for all of their repository objects. When all respondents are taken into account however, nearly a third of repositories (fifteen, or 31 percent) record no technical metadata.



[insert figure 8]

Respondents were then asked to gauge their satisfaction with the amount of technical metadata that can be collected and the amount that actually is collected. Similar to the results with rights metadata, repository capability (“technical metadata you CAN collect”) had a higher rating average (2.43) than actual technical metadata practice (“technical metadata you DO collect”) (1.95).

Forty respondents assessed the amount of technical metadata the repository can accommodate. Of those, sixteen (40 percent) felt the amount of metadata accommodated by their schema is “just right.” In terms of actual practice, only ten (25 percent) felt the amount of metadata they routinely collect is “just right”; 35 percent felt the amount was “way too little.”

[insert figure 9]

Survey responses were again reviewed one-by-one to determine frequency and size of any gaps between technical metadata accommodated and technical metadata routinely recorded. For 69 percent of respondents, there was no gap.<sup>46</sup> Nine percent had a gap of one rating point; 20 percent had a gap of two rating points; 3 percent had a gap of three rating points.

Finally, data were analyzed to determine if the rating for a repository’s technical metadata *capability* correlated to the number of metadata elements the schema accommodates (as derived from question 16). Respondents were divided into groups of roughly equal size, based on the number of metadata elements accommodated, and ratings were compared for each category.

[insert table 4]

In general, the more metadata elements a schema accommodates, the more likely respondents are to rate the metadata they *can* collect as 3 (“just right”) or 4 (somewhere between “just right” and “way too much”).

[insert table 5]

However, there appears to be no consensus as to how many elements would warrant a “just right” rating. For the twenty-one respondents who rated the repository capability “just right,”<sup>47</sup> the number of elements ranged from just one at one end of the spectrum to ten at the other end.

[insert table 6]

### **Preservation metadata**

Just over half of the surveyed repositories<sup>48</sup> (28 of 54, or 52 percent) accommodate some preservation metadata.

Respondents were presented with a list of thirteen preservation elements and asked to check those accommodated by their repository metadata schema.<sup>49</sup> Every element offered on the list was used by at least one repository. The average number of preservation elements accommodated was 3.5. The three most common are:

- 1) Storage location (file location; location scheme, e.g., handle, URI; storage medium, e.g., hard disc, magnetic tape, etc.) (63 percent)
- 2) Links between objects when one is derived from the other (37 percent)
- 3) Preservation level (bit-level, full, etc.) (33 percent)

[insert figure 10]

Respondents were asked the portion of the repository's objects for which preservation metadata is actually recorded. Of the twenty-six respondents whose repositories accommodate some preservation metadata and who were able to answer this question, twenty-one (81 percent) record some of this preservation metadata for all their repository objects.

Given that three of the respondents presented with this question do not actually record any of this preservation metadata, it can be said that of all fifty-four survey respondents, exactly half record no preservation metadata at all.

[insert figure 11]

Respondents were next asked to gauge their satisfaction with the amount of preservation metadata that can be collected and the amount that actually is collected. Preservation metadata capability ("preservation metadata you CAN collect") had a higher rating average (2.48) than actual preservation metadata practice ("preservation metadata you DO (routinely) collect") (2.0).

Twenty-five respondents assessed the amount of preservation metadata the repository can accommodate. Of those, eight (32 percent) felt the amount of metadata accommodated by their schema is "just right".

In terms of actual practice, twenty-seven respondents made the assessment and eight of those (30 percent) felt the amount of metadata they routinely collect is "just right."

[insert figure 12]

Survey responses were again reviewed individually to determine frequency and size of any gaps between metadata accommodated and metadata routinely recorded. For 68 percent of respondents, there was no gap; 4 percent had a gap of one rating point; 11 percent had a gap of two rating points; and 4 percent had a gap of four rating points.

Finally, data were analyzed to determine if the rating for a repository's preservation metadata *capability* correlated to the number of metadata elements the schema accommodates (as derived from question 23). Respondents were divided into groups of roughly equal size, based on the number of metadata elements accommodated, and ratings were compared by category.

[insert table 7]

In general, the more metadata elements a schema accommodates, the more likely respondents are to rate the metadata they *can* collect between 3 (“just right”), and 5 (“way too much”).

[insert table 8]

As before, however, there appears to be no consensus as to how many elements would warrant a “just right” rating. The number of elements rated “just right” (3) ranged from one to 11; those rated somewhat below that (2) were in a similar range.

[insert table 9]

### **Source metadata**

Respondents whose repositories accommodate source metadata were asked only the broad categories of source metadata accommodated (physical characteristics, provenance, physical location), and the percentage of applicable resources to which the source metadata was actually applied. Over half of surveyed repositories (29 of 54, or 54 percent) accommodate some source metadata.

Respondents were asked which general types of source metadata they record. The most common information recorded was that pertaining to physical characteristics (93 percent), followed by provenance (70 percent), then physical location (59 percent).<sup>50</sup> Fewer than half (48 percent) recorded all three types.

Just over a quarter (26 percent) recorded just two types, and an equal number (26 percent) recorded just one type.

Nearly a third of respondents (eight, or 31 percent) record some source metadata for all of their source objects; 23 percent record source metadata for less than one third of their source objects, and of those, 4 percent record no source metadata.<sup>51</sup>

If all fifty-four survey respondents are taken into account, at least 48 percent record no source metadata, although at least one (who stated this explicitly), and possibly others, hold only born-digital content and would have no source metadata to record.

[insert figure 13]

Next, respondents were asked to gauge their satisfaction with the amount of source metadata that can be collected and the amount that actually is collected. Source metadata capability (“source metadata you CAN collect”) had a higher rating average (2.74) than actual source metadata practice (“source metadata you DO (routinely) collect”) (2.48).

Twenty-seven respondents assessed the amount of source metadata the repository can accommodate. Of those, nineteen (70 percent) felt the amount of metadata accommodated by their schema is “just right.” In terms of actual practice, only fifteen (56 percent) felt the amount of metadata they routinely collect is “just right.”

[insert figure 14]

Survey responses were again reviewed individually to determine frequency and size of any gaps between metadata accommodated and metadata routinely recorded. For 80 percent of respondents, there was no gap; 16 percent had a gap of one rating point; 4 percent had a gap of two rating points.

Repository tasks supported by repository metadata

Respondents were offered a list of nineteen repository tasks, which have been established in the literature<sup>52</sup> as key functions of trusted repositories (although it is interesting to note that not all respondents characterized their repositories as preservation repositories). Respondents were then asked which of the tasks is supported by the repository's metadata. Of fifty-three respondents,<sup>53</sup> the average number of tasks is 7.1. The four most common tasks are:

- 1) Store original digital object
- 2) Document owner of intellectual property rights
- 3) Protect against data corruption and loss
- 4) Track origins and chain of custody of digital object (provenance)

[insert figure 15]

Any one task pertains to a particular repository function supported by one or more types of metadata, as follows. Because some functions are supported by more than one kind of metadata, some tasks are assigned to more than one category.

Object Use Tasks (supported by rights metadata; cf. Question 12)

- Document owner of intellectual property rights
- Document permissions to distribute, duplicate, transfer, and/or alter (e.g., through migration)
- Store contractual agreements pertaining to rights and permissions
- Enable repurposing of content (e.g., for revenue generation)

Object Fixity, Integrity, and Authenticity Tasks (supported by technical metadata; cf. Question 16)

- Store original digital object
- Protect against data corruption and loss
- Ensure authenticity of digital resources over time
- Track origins and chain of custody of digital object (provenance)

- Document digital object's bitstream over the long term
- Ensure resources will survive and continue to be understandable into the long term
- Maintain vigorous and ongoing testing and validation program to ensure independent understandability of the digital object
- Enable format migration/transformation upon obsolescence
- Track migration path of digital object and any changes over time (digital provenance)
- Document relationships between multiple manifestations of a digital object

Object Preservation Tasks (supported by technical, preservation, and source metadata; cf. Questions 16, 23, 27)

- Store original digital object
- Protect against data corruption and loss
- Ensure authenticity of digital resources over time
- Track origins and chain of custody of digital object (provenance)
- Document digital object's bitstream over the long term
- Ensure resources will survive and continue to be understandable into the long term
- Maintain vigorous and ongoing testing and validation program to ensure independent understandability of the digital object
- Enable format migration/transformation upon obsolescence
- Track migration path of digital object and any changes over time (digital provenance)
- Document relationships between multiple manifestations of a digital object
- Facilitate condition assessment of digital preservation master
- Facilitate decision-making for preservation managers
- Document preservation actions taken
- Document effects of preservation strategies
- Document details of original source object (e.g., provenance, preservation, condition)

Logically, the richer a repository's metadata in any given area, the more tasks in that area it can support. To test this hypothesis, the number of tasks (perceived to be) supported by each repository was charted alongside the number of metadata elements accommodated by that repository. For each task/metadata type pair, the number of tasks supported tends to rise with the number of metadata elements accommodated, as shown below.

[insert table 10]

[insert table 11]

[insert table 12]

Taking all respondents into account, repository metadata supported, on average, 1.6 of four use-related tasks (40 percent), 4.5 of ten object fixity/integrity/authenticity tasks (45 percent), and 5.8 of fifteen preservation tasks (39 percent).

### **Additional factors impacting the recording of metadata creation**

Much of the data above show that in general, metadata accommodated and recorded is insufficient to carry out many repository tasks. Several factors may account for the shortfall. One obvious factor could be a lack of staff resources. Use of automated metadata tools could mitigate the problem, at least insofar as technical metadata is concerned. Therefore, the survey included three questions on staffing and metadata extraction tools.

Respondents were first asked who creates their administrative metadata; the most common responses were: Digital Repository Department, Depositor, System supplied defaults, and Cataloging and Metadata Department. On average, administrative metadata emanated from an average of 2.8 sources.



Next, respondents were asked if the institution's staffing was sufficient to perform the tasks they had said were supported by their repository metadata; 45 percent said yes and 37 percent said no.

As for metadata extraction tools, only ten surveyed repositories use metadata extraction tools (23 percent of the forty-three accommodating technical metadata). The most commonly used tool is the JSTOR/Harvard Object Validation Environment (JHOVE) (nine), followed by Exiftool, the National Library of New Zealand's metadata extraction tool, and Flexible Image Transport System (FITS).

## **Conclusion**

This study surveyed fifty-four ARL institutional repositories about their administrative metadata capabilities and practices. Responses throughout the survey indicate that in general, organizations are not accommodating administrative metadata to any significant extent. For example, on average, repositories accommodate only 3.6 rights metadata elements, three technical metadata elements, and 3.5 preservation elements. While organizations may record more metadata than the numbers suggest (since these numbers reflect only the number of elements dedicated to a particular piece of metadata), the parsing of that metadata is not sufficient to enable efficient retrieval, machine processing, reporting, and sharing. In any case, nearly a quarter (23 percent) record no rights metadata; nearly a third (31 percent) record no technical metadata; and exactly half record no preservation metadata. Of all possible elements that might be accommodated by the repositories' schemas, only three were in use by more than half the surveyed repositories (format, file size, and rights statement). Only 35 percent of the surveyed repositories had a dedicated metadata element for storage location, which (interestingly) was the most commonly accommodated preservation metadata.

Moreover, across the board, for all metadata types, repository capability (measured by number of dedicated administrative metadata elements accommodated), outranks actual practice. For example, of those accommodating *some* rights metadata, less than a third recorded "some of" the rights metadata offered in the survey questions all of the time. Not surprisingly, few respondents think the amount of

metadata they actually record is “just right” (between 32 percent and 42 percent, depending on metadata type). Responses suggest that in nearly all cases, where the metadata is not “just right,” it falls short.

Administrative metadata cannot be assessed outside the context of the functions trusted repositories are meant to perform. Administrative metadata supports repository tasks related to object fixity, integrity, authenticity, use, and preservation. Without it, those tasks cannot be carried out. This survey found that of nineteen key repository tasks identified in the literature, an average of only 7.1 are supported by the repositories’ metadata. Measured in terms of the metadata type supporting each task, repositories perform better, percentage-wise, on the fixity/integrity/authenticity tasks, and are less prepared to support object use and preservation tasks.

There are many possible reasons for these “metadata shortfalls.” The influx of digital materials is rapid and increasing at a rate that may well put their management beyond the means of most institutions. Metadata standards are voluminous and complex, and repositories must employ multiple standards to cover the necessary range of administrative metadata, complicating implementation within any one system. Staffing may be an issue. Most surveyed institutions spread administrative metadata work across multiple units, implying a diversity of workflows and reporting structures. Thirty-seven percent of respondents found their staffing insufficient to perform repository tasks. Perhaps more discussion is needed to convey the importance of administrative metadata to digital collections management, preservation, and use. At the institutional level, it is unclear from this study who is determining what metadata will be accommodated and recorded, and whether those decisions are based on established repository objectives.

In any case, collaborative preservation is often assumed to be a universal good,<sup>54</sup> but if these survey results are an indicator, preservation of any kind appears to be a local phenomenon. The community is not putting enough effort into administrative metadata and the paucity of metadata being collected cannot support collaborative preservation. There seems to be a significant disconnect between what the community is saying and what is actually happening on the ground. There may even be skepticism in the community that collaborative preservation is valuable and possible.

This study points to several areas for further research. First, it is critically important to identify the core administrative metadata required to effectively manage and preserve repository resources. The best way to identify core metadata is to enumerate required repository tasks, then determine which individual pieces of metadata support each task. Once that work is done, libraries would benefit from exploring other obstacles to collecting administrative metadata. These might include issues related to self deposits, legacy metadata, expertise, and staffing levels. Any study that thoughtfully explores obstacles to the collection of rich, or even core, administrative metadata, could make a significant contribution to the field.

Reaching consensus on core administrative metadata is central to resolving current preservation challenges. If libraries are to make good on their promise to provide permanent, organized and secure repositories for institutional scholarship and special collections, they will need to examine current practices and identify core metadata in the context of repository objectives. They must identify barriers to metadata collection, strategize as to how those barriers might be mitigated or overcome, and move forward.

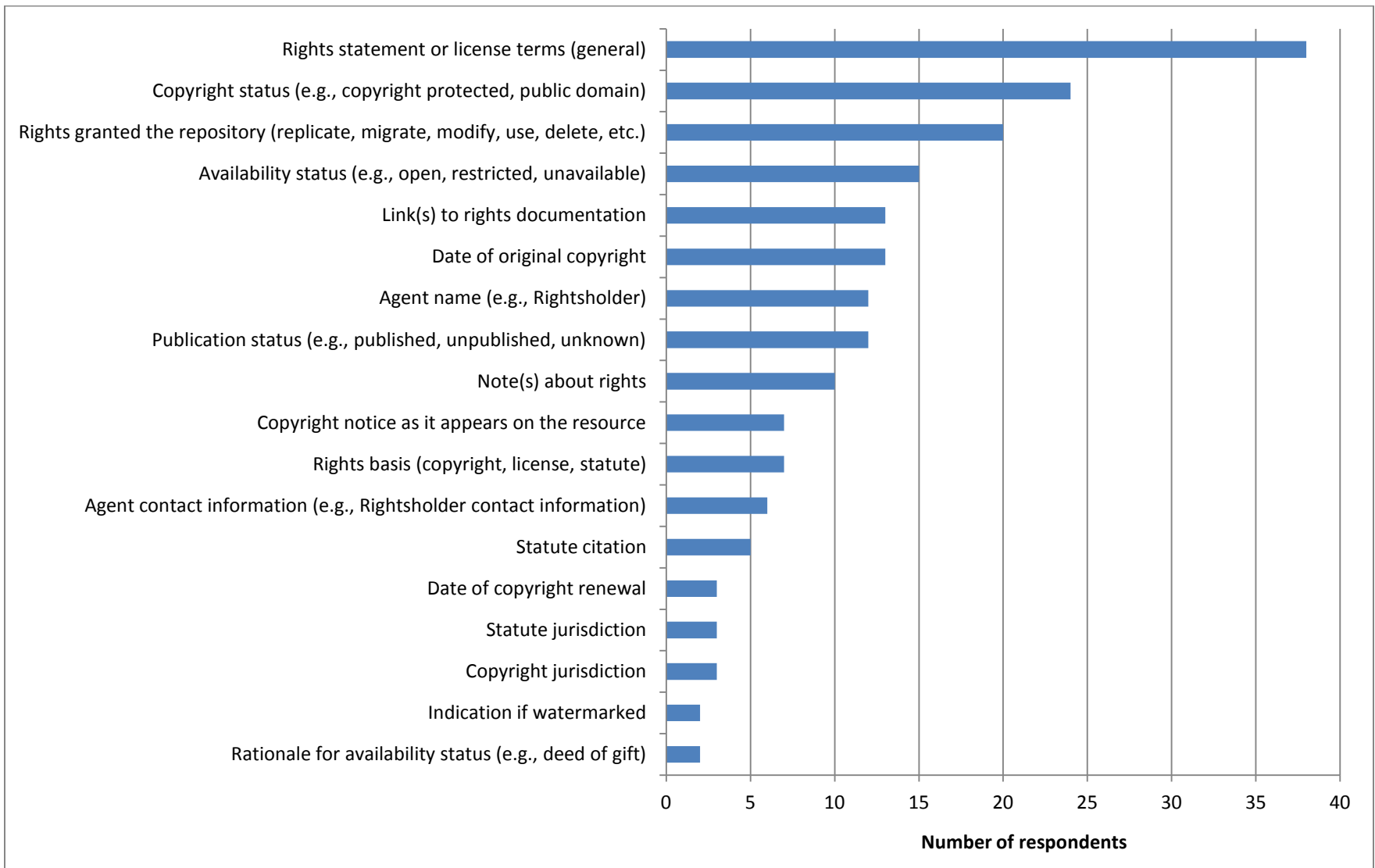


Figure 1. Rights metadata elements accommodated

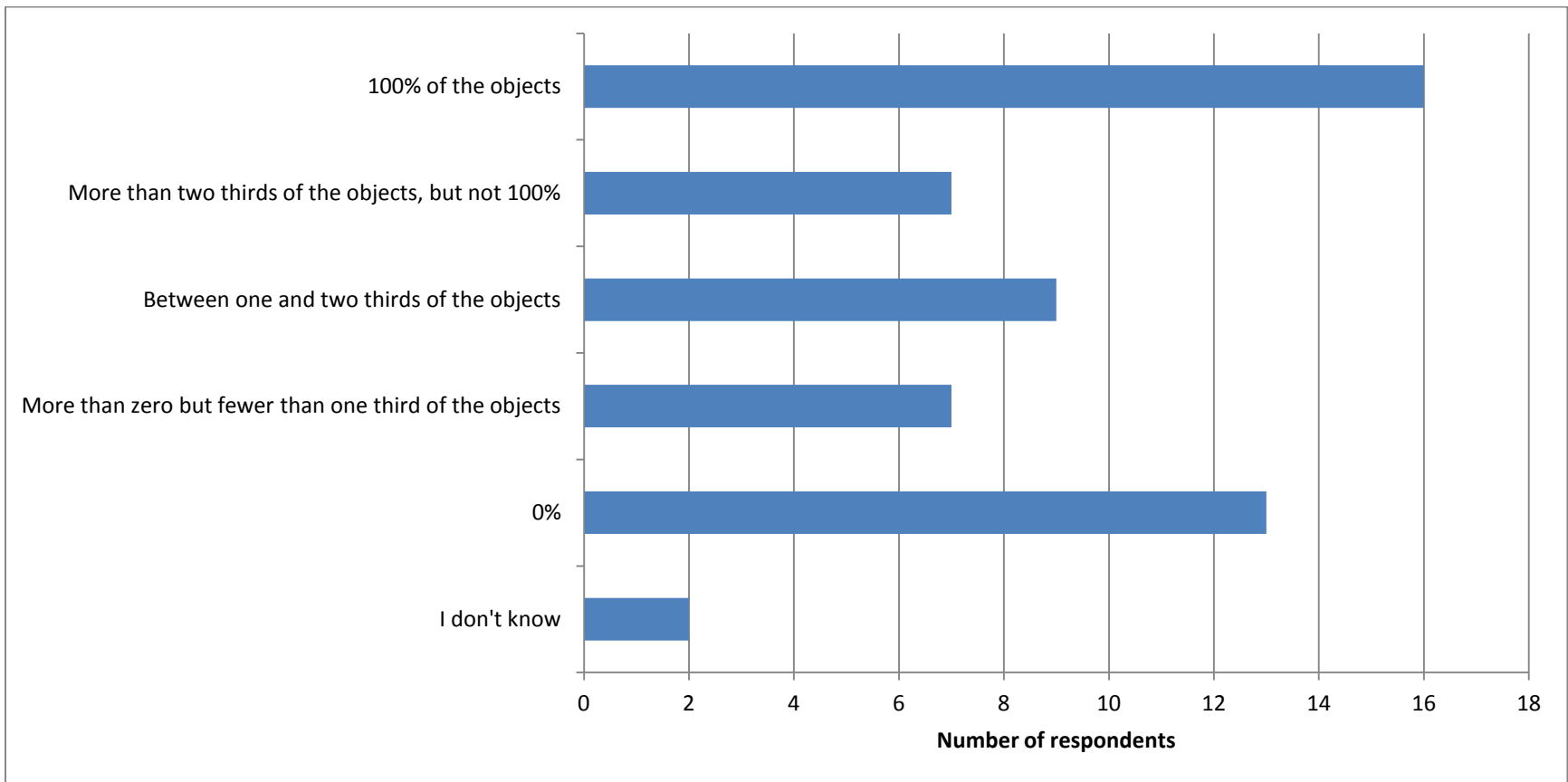


Figure 2. For what portion of your repository objects (roughly) do you record some of this rights metadata? (Includes repositories not accommodating rights metadata.)

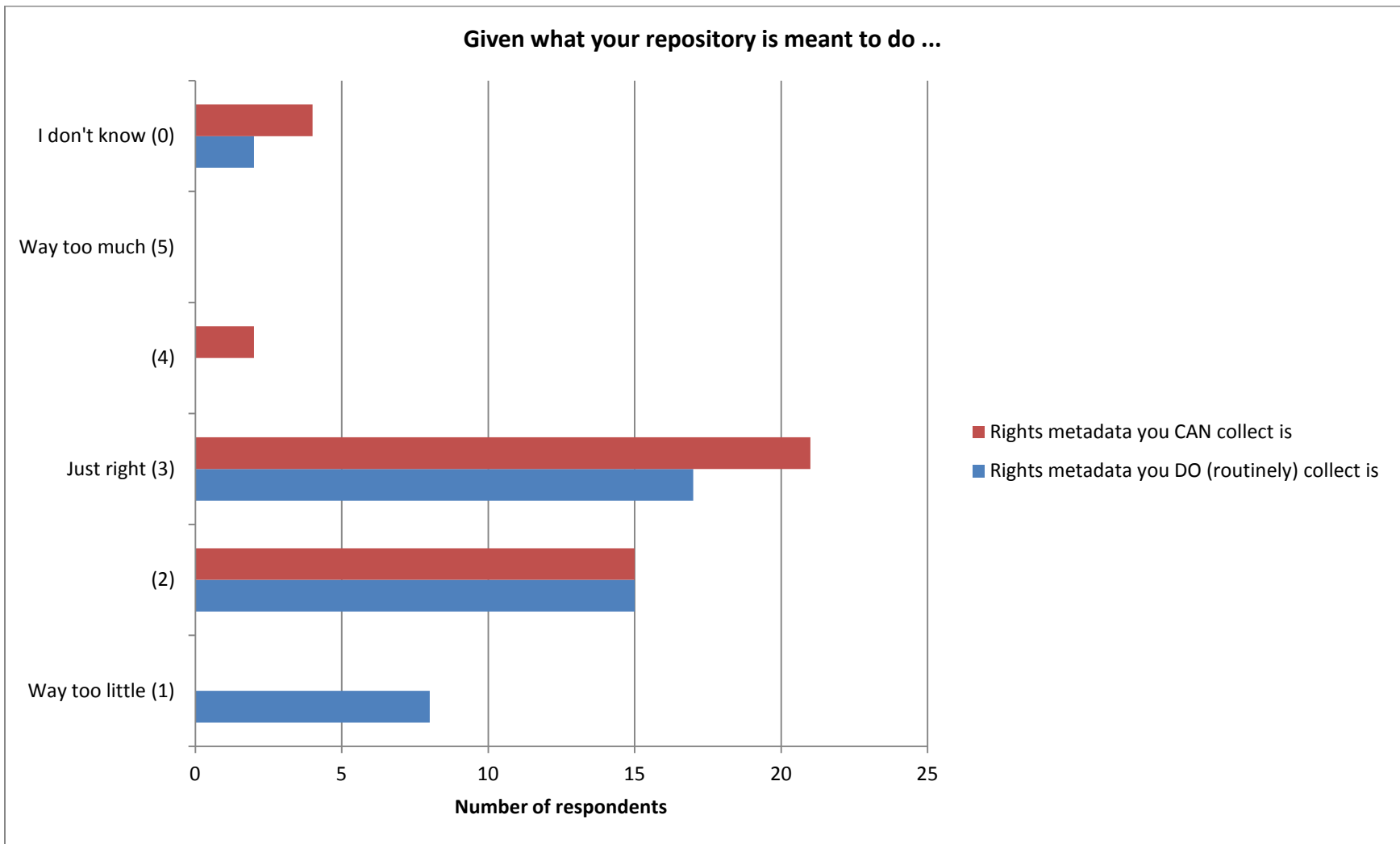


Figure 3. Rights metadata capability compared to rights metadata practice

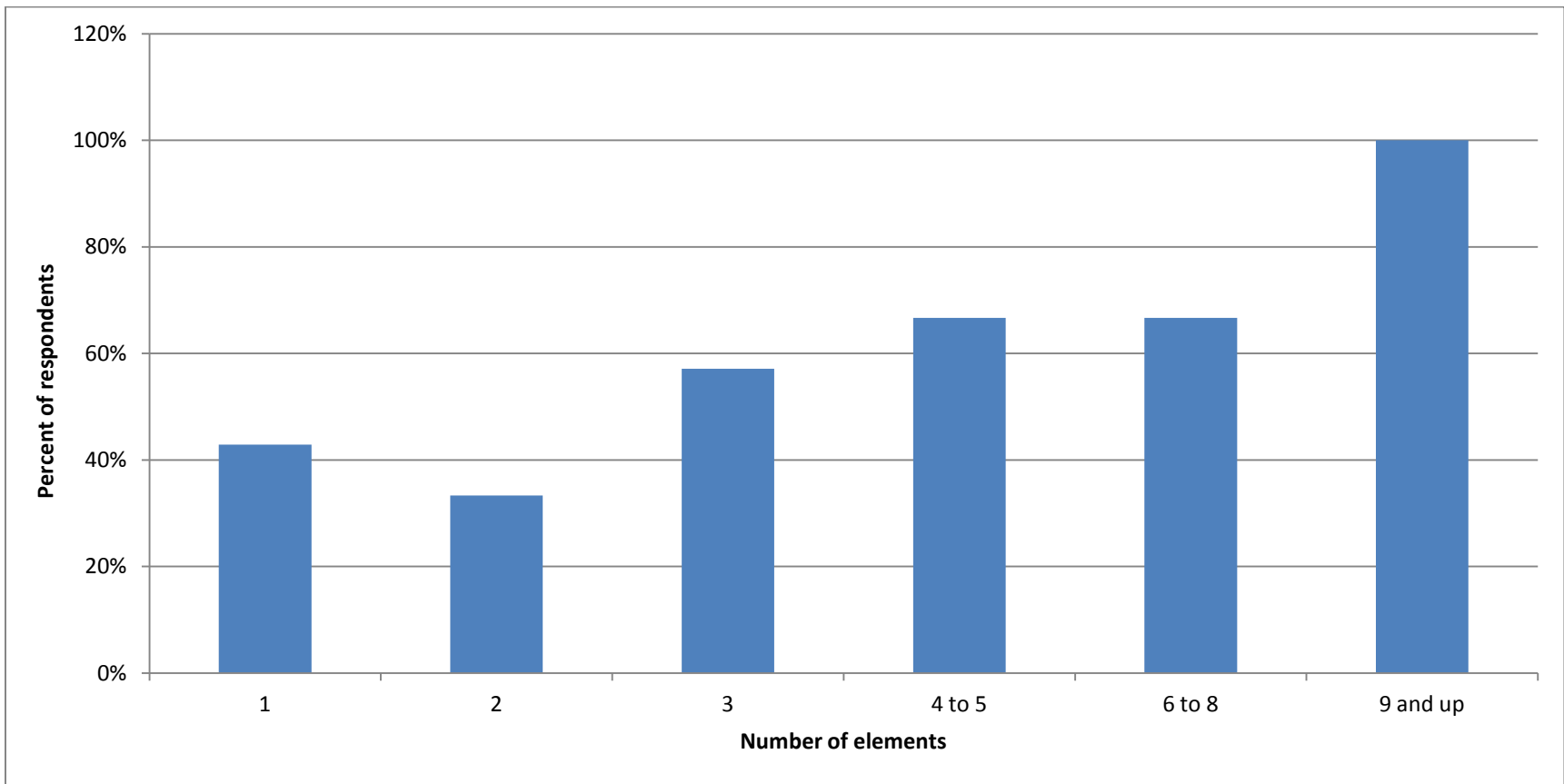


Figure 4. Percent of respondents rating rights metadata 3 ('just right') or higher (1=way too little; 3=just right; 5=way too much)

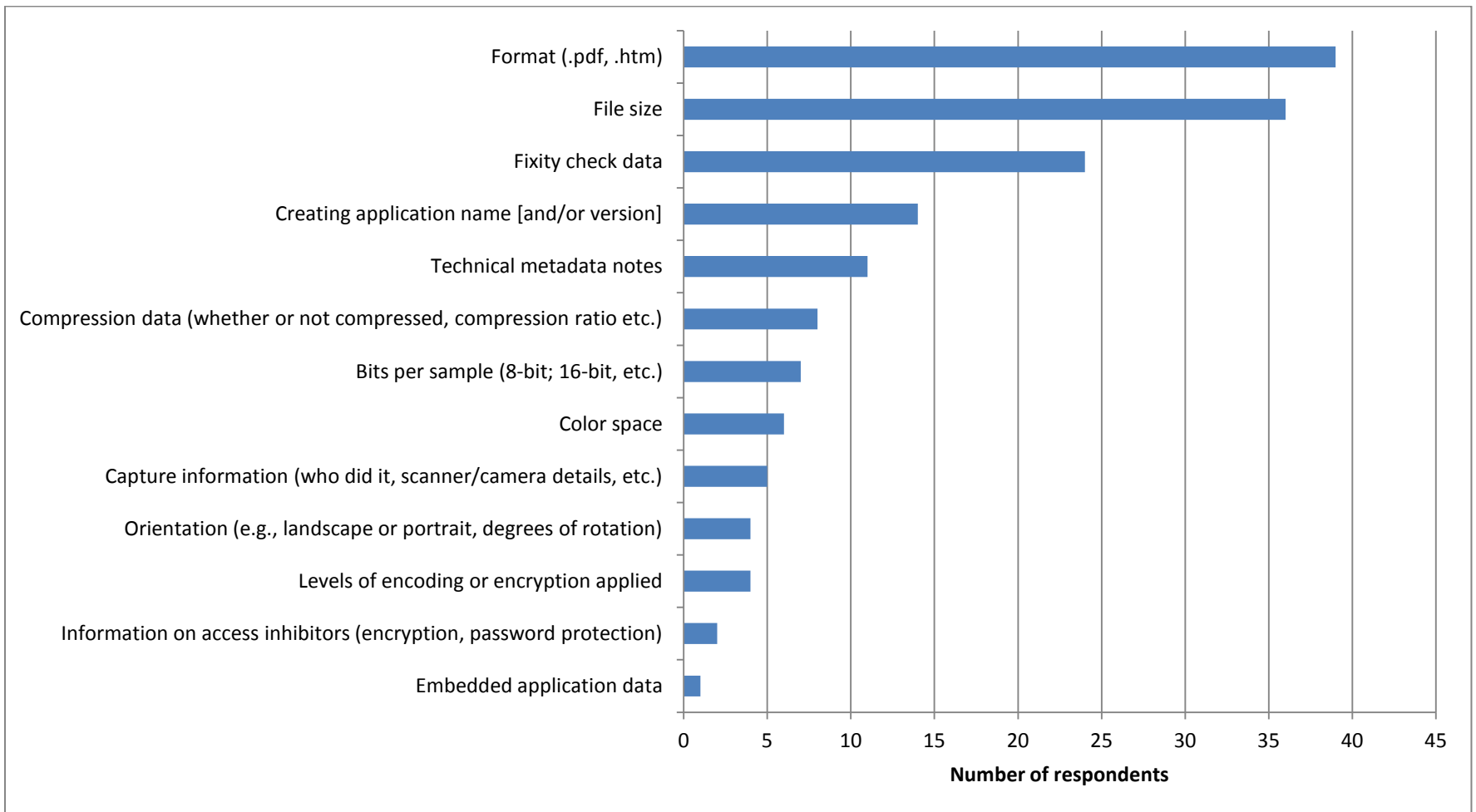


Figure 5. Technical metadata elements accommodated



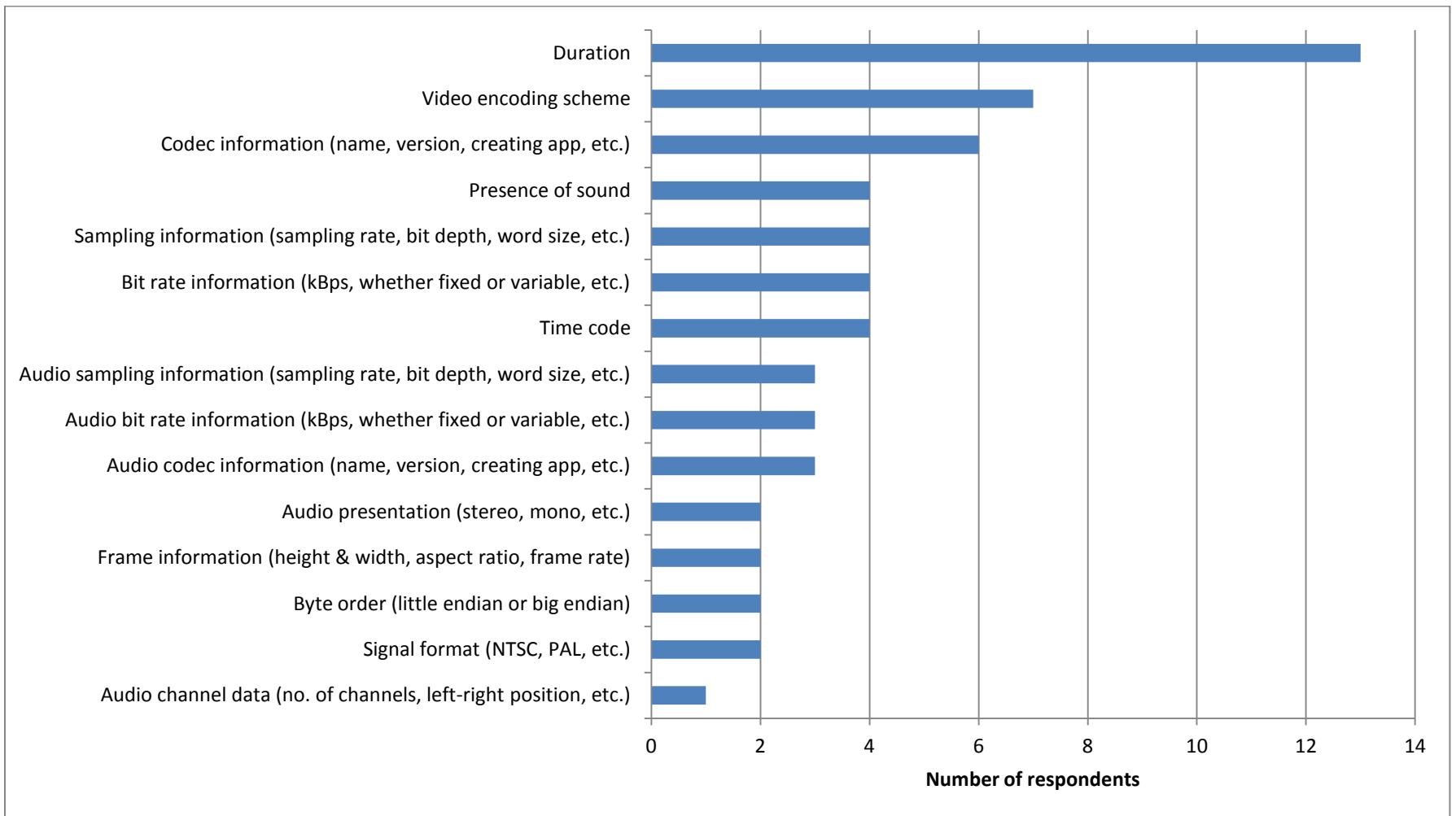


Figure 6. Technical metadata elements for video accommodated

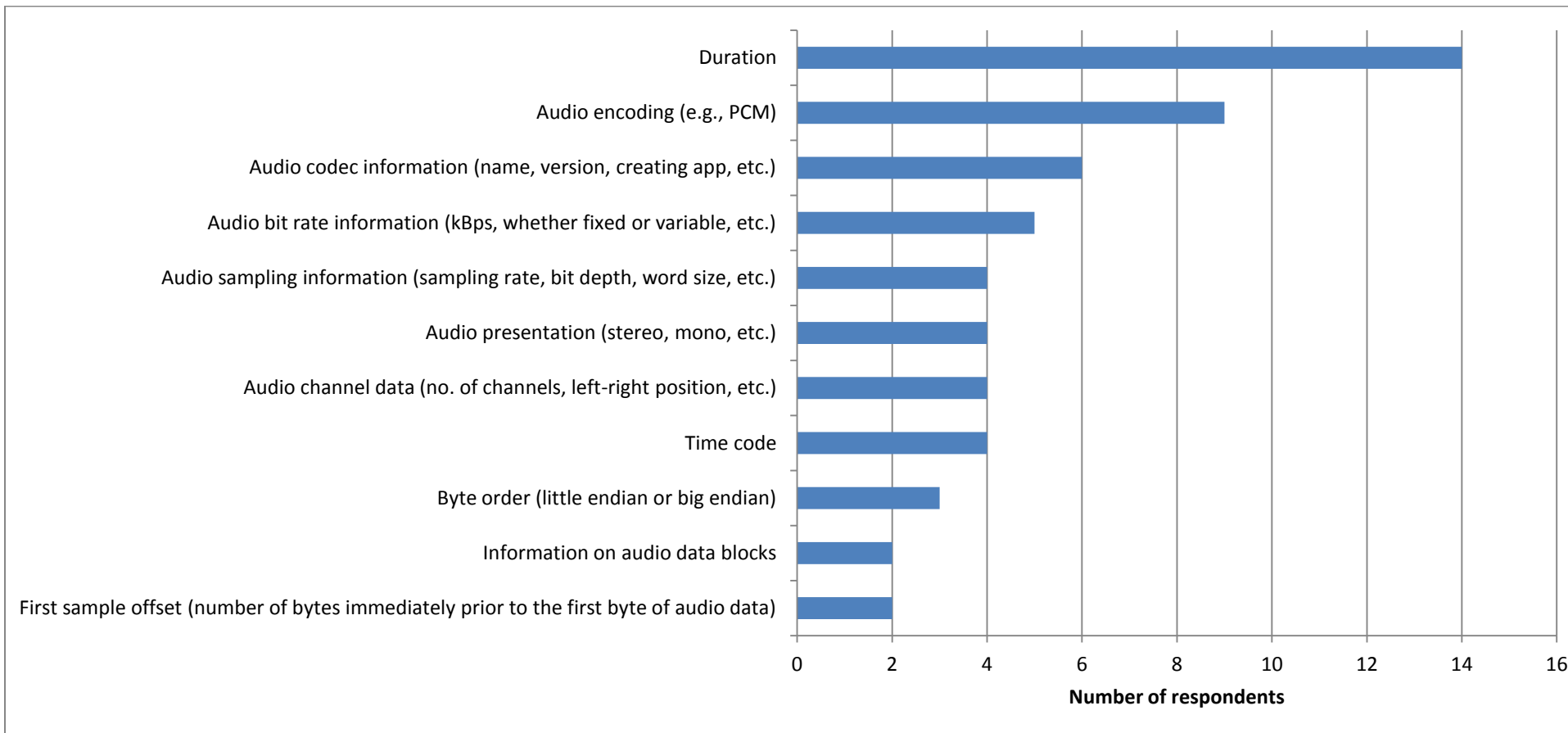


Figure 7. Technical metadata elements for audio accommodated

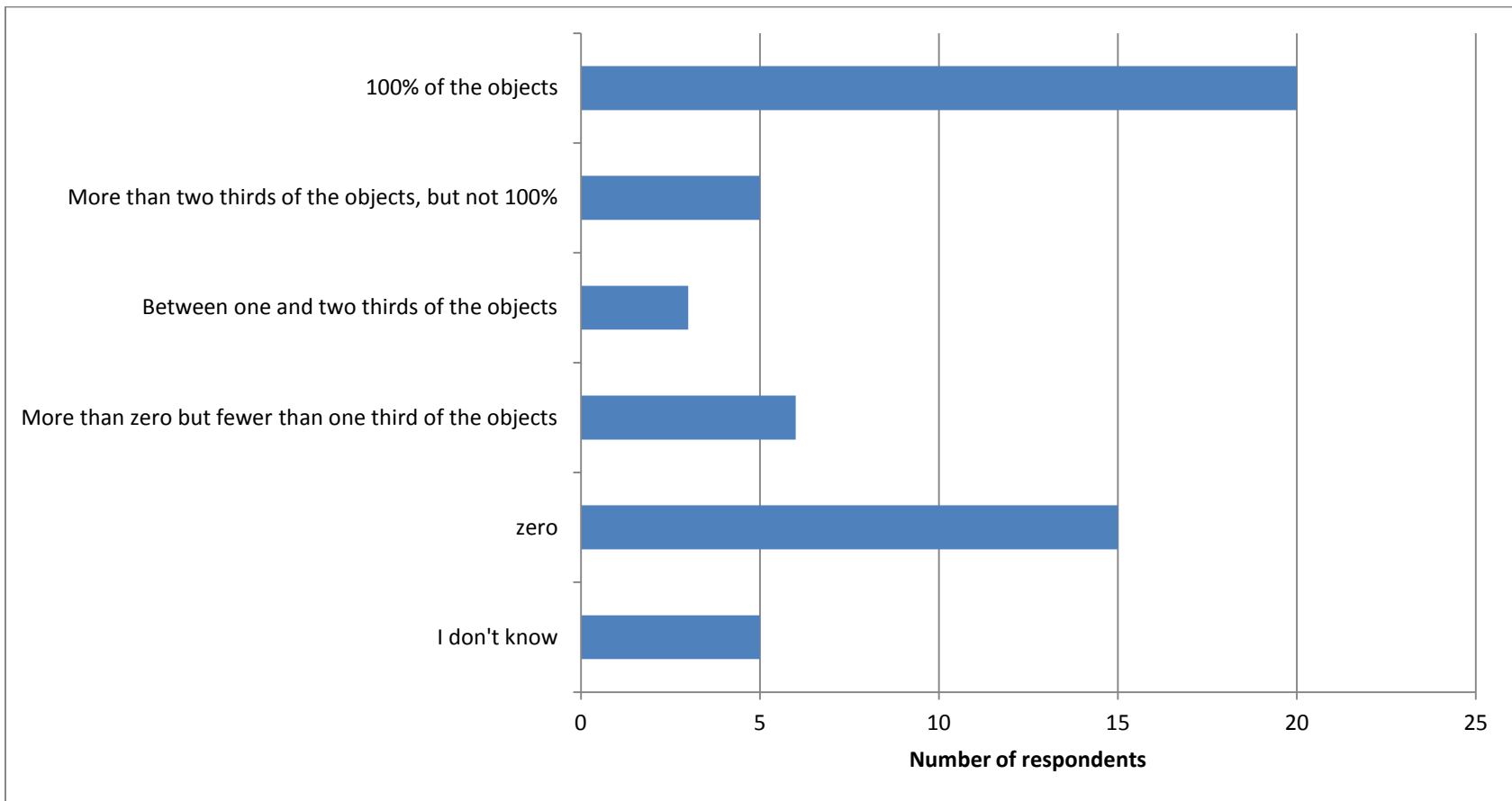


Figure 8. Portion of repository objects for which some of this technical metadata is recorded (Includes repositories not accommodating technical metadata)

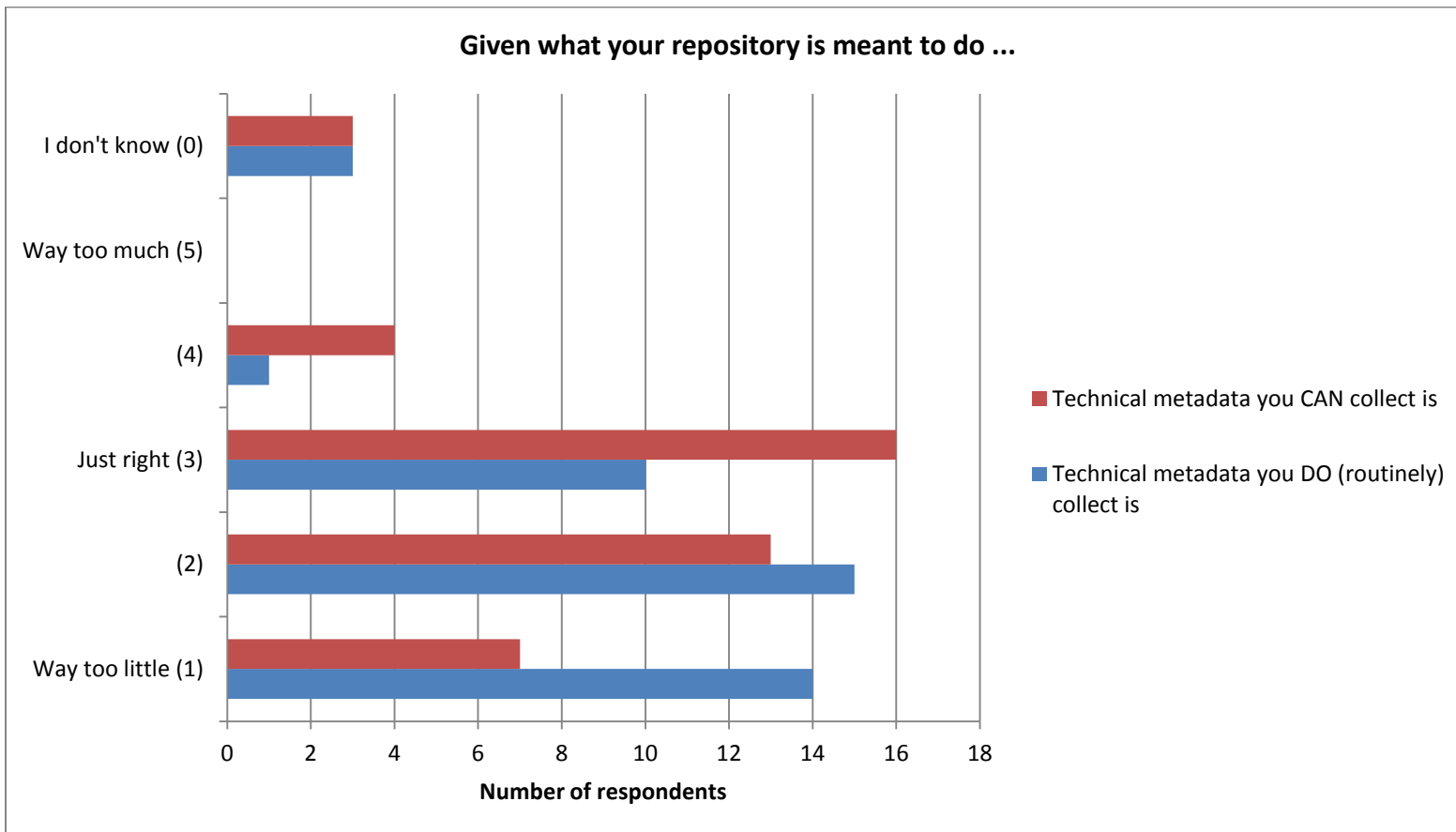


Figure 9. Technical metadata capability compared to technical metadata practice

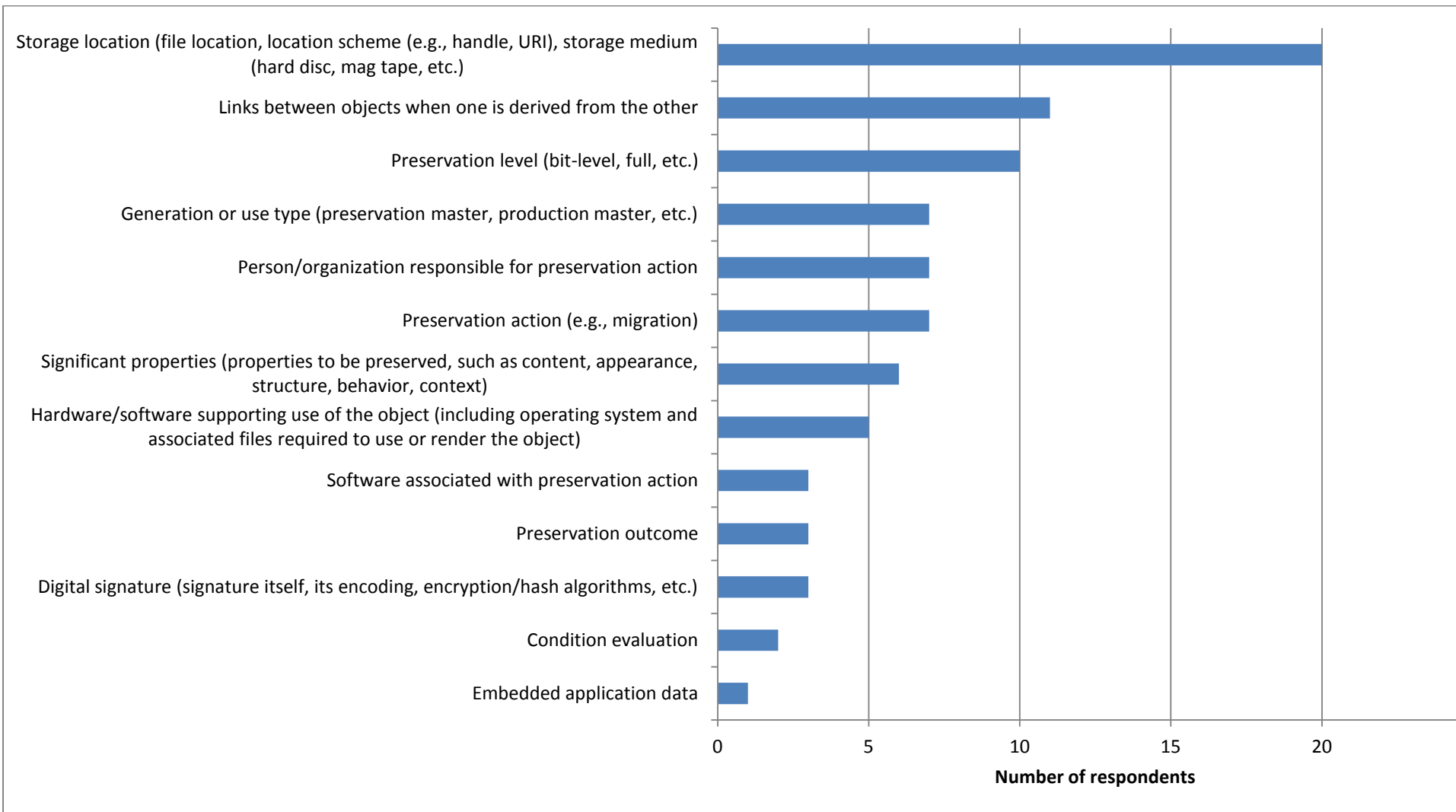


Figure 10. Preservation metadata elements accommodated

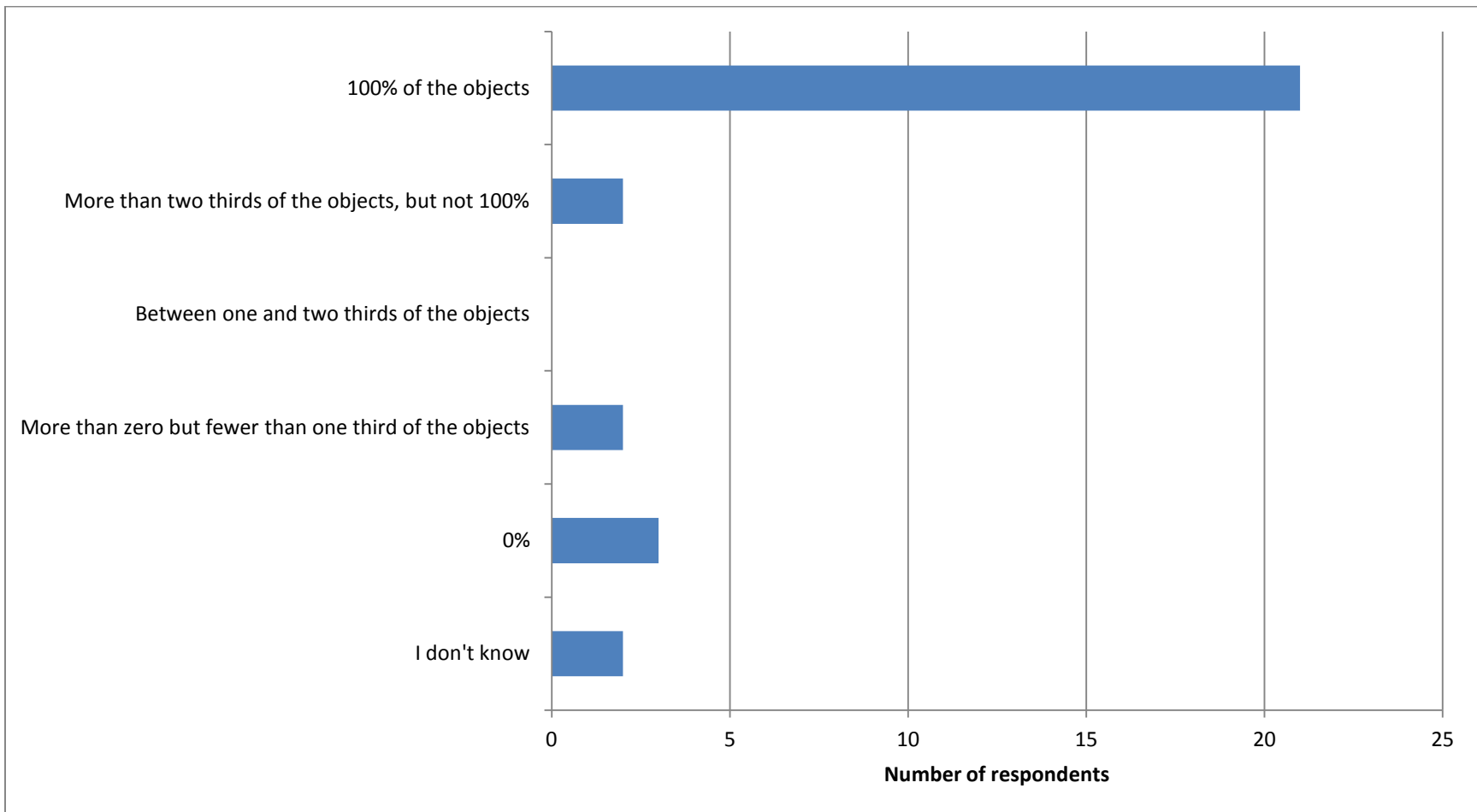


Figure 11. Portion of repository objects (roughly) for which some of this preservation metadata is recorded (includes repositories not accommodating preservation metadata)

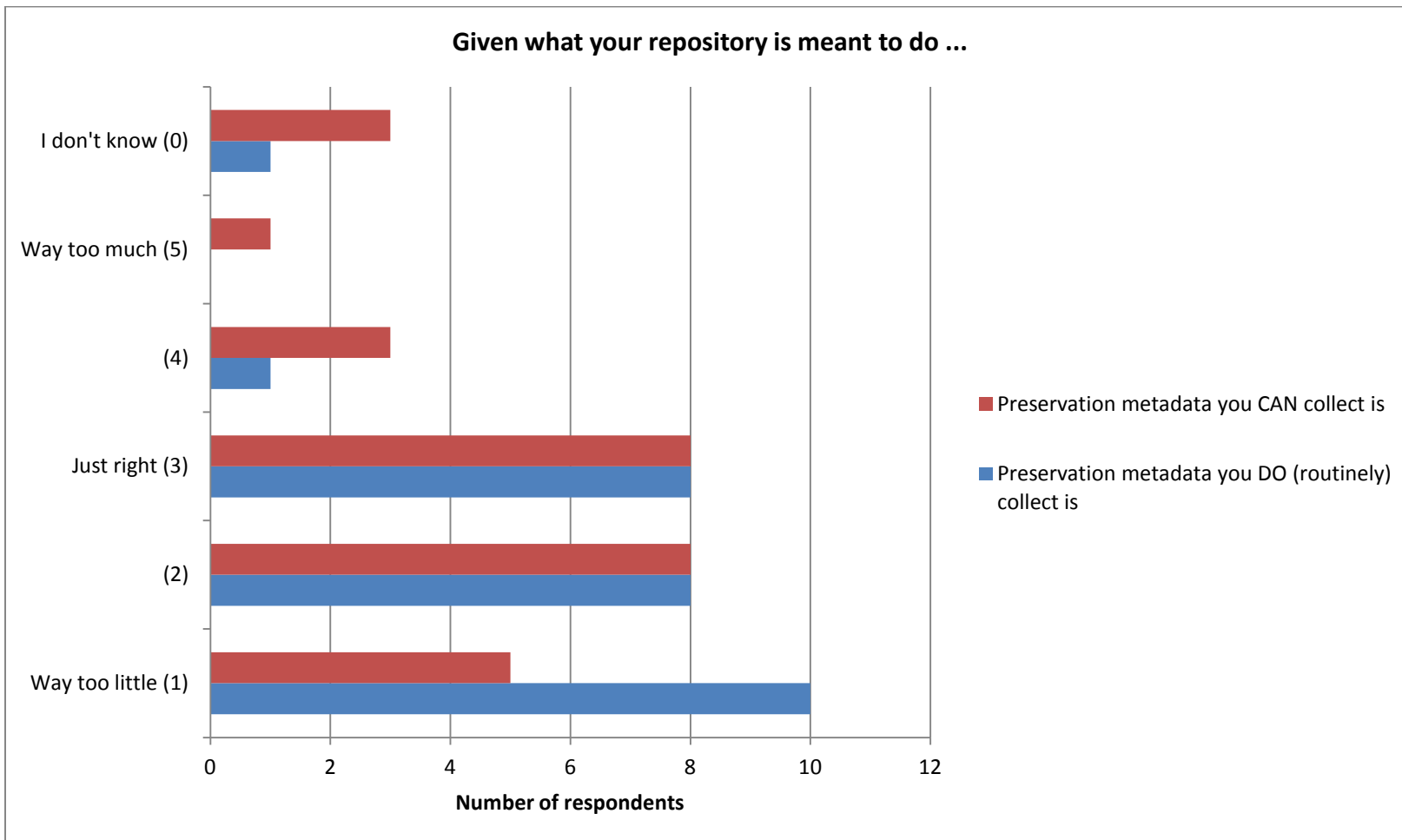


Figure 12. Preservation metadata capability compared to preservation metadata practice

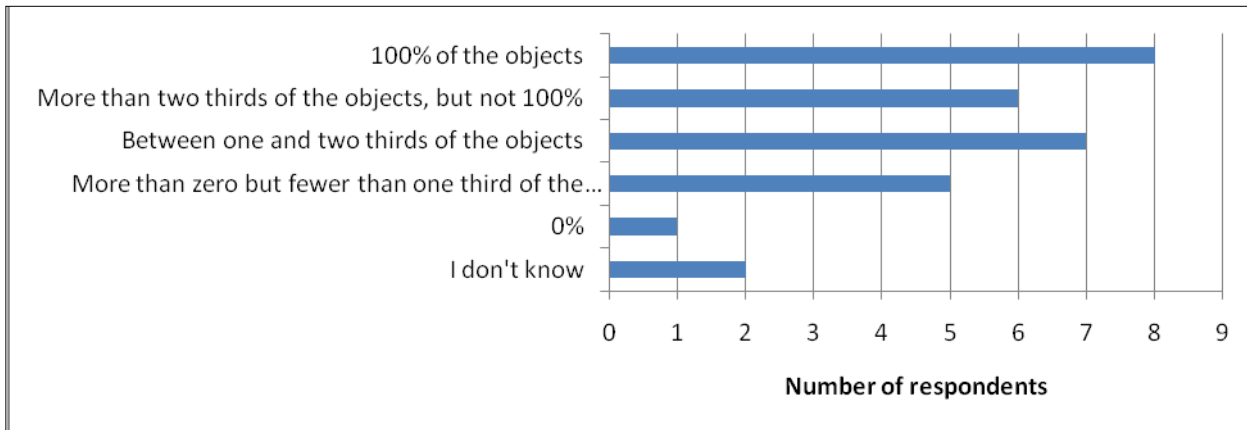


Figure 13. Portion of repository source objects for which some source metadata is recorded



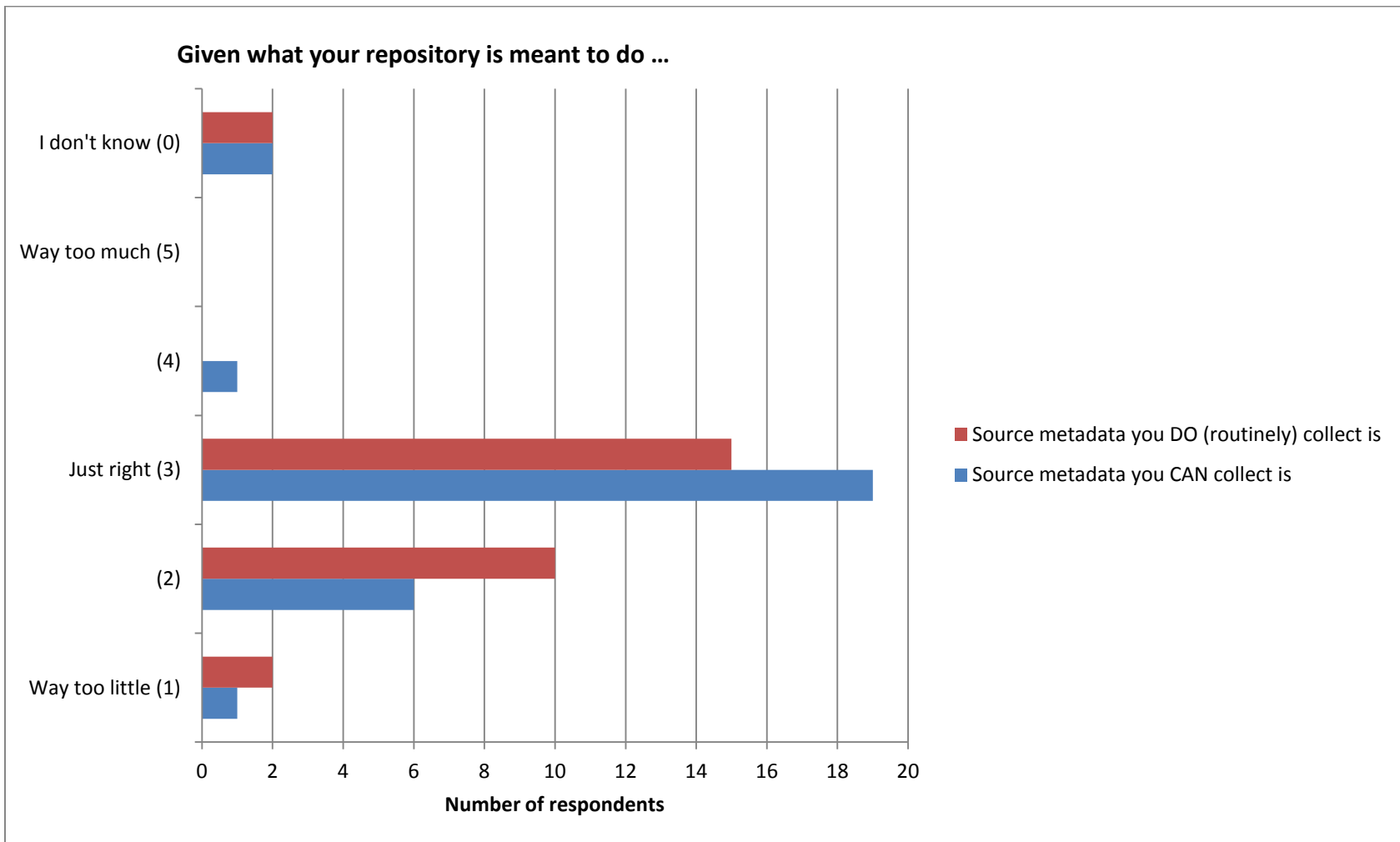


Figure 14. Source metadata capability compared to source metadata practice

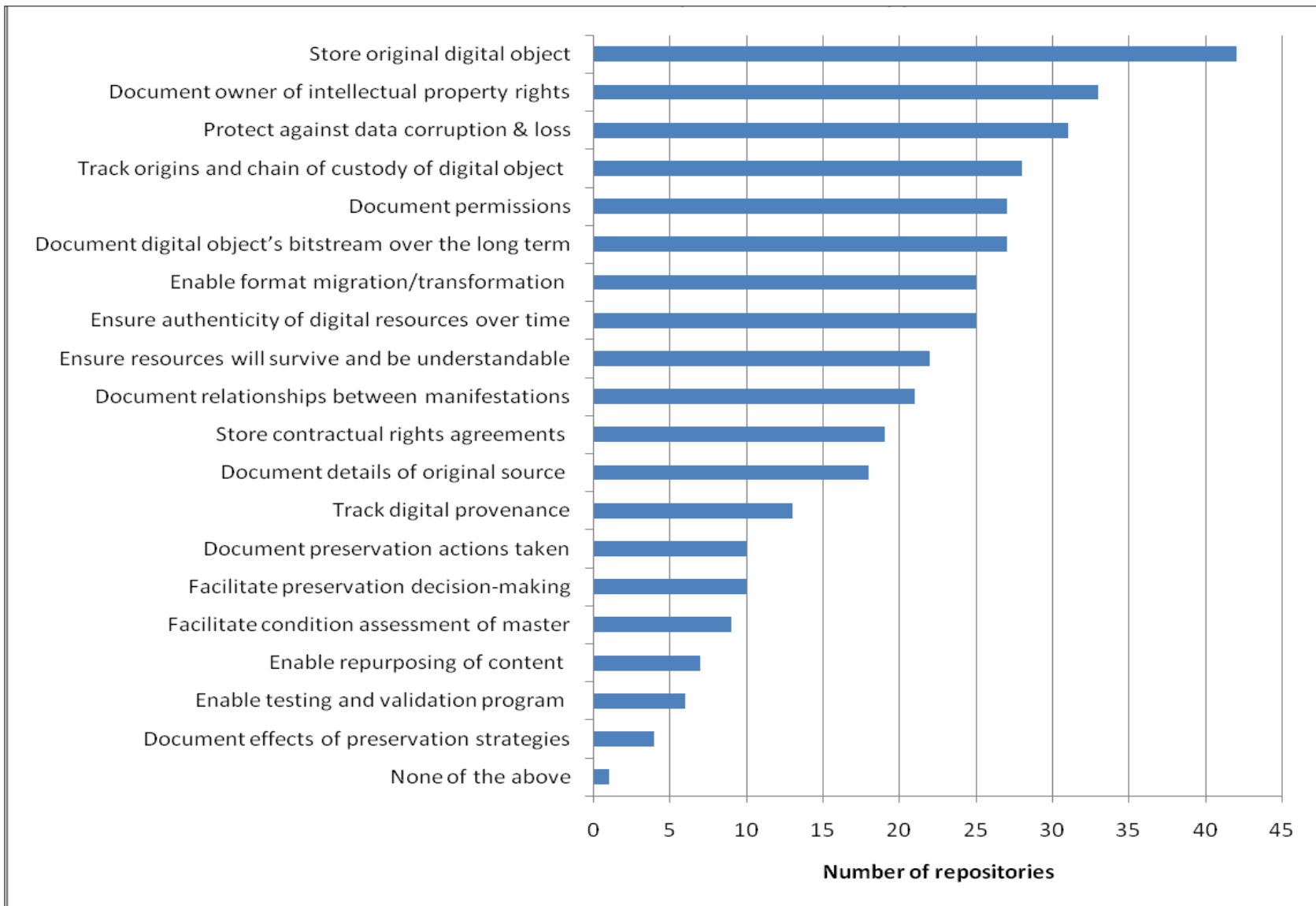


Figure 15. Tasks supported by repository metadata

Rank	Standard	Number of respondents
1	Qualified Dublin Core	43
2	Simple Dublin Core	23
3	MODS	16
4	PREMIS	11
5	NISO MIX (Z39.87)	10
6	MARC	9
	Other	23

Table 1. Administrative metadata standards incorporated into repository metadata

<b>Rights metadata you <i>can</i> collect is ...</b>						
No. of elements accommodated by the repository	1 “way too little”	2	3 “just right”	4	5 “way too much”	0 “don’t know”
1 (7 respondents)	0	4	3	0	0	0
2 (6 respondents)	0	4	2	0	0	0
3 (7 respondents)	0	5	2	0	0	1
4-5 (6 respondents)	0	2	4	0	0	2
6-8 (6 respondents)	0	2	4	0	0	0
9 + (6 respondents)	0	0	5	1	0	2
Total	0	17	20	1	0	5

Table 2. Ratings of rights metadata capability

<b>Rights metadata you <i>can</i> collect is ...</b>			
Rating	No. of respondents with this rating	No. of elements accommodated by repositories with this rating	Average no. of elements accommodated by repository with this rating
1 (way too little)	0	n/a	n/a
2	15	1-8	2.87
3 (just right)	21	1-18	5.57
4	2	3 and 13	8
5 (way too much)	0	n/a	n/a
Don't know	4	3-12	6.25

Table 3. Rating of rights metadata capability relative to number of rights elements accommodated

<b>Technical metadata you <i>can</i> collect is ...</b>						
No. of elements accommodated by the repository	1 “way too little”	2	3 “just right”	4	5 “way too much”	0 “don’t know”
1-2 (12 respondents)	4	3	4	0	0	1
3 (12 respondents)	2	4	4	1	0	1
4 (8 respondents)	1	3	4	0	0	0
5-11 (9 respondents)	0	2	3	3	0	1
Total	7	12	15	4	0	3

Table 4. Ratings of technical metadata capability

Number of elements	Number of respondents rating 3 or higher	Percent of respondents rating 3 or higher
1-2 (12 respondents)	4	33 percent
3 (12 respondents)	5	42 percent
4 (8 respondents)	4	50 percent
5-11 (9 respondents)	6	66 percent

Table 5. Percent of respondents rating technical metadata 3 ('just right') or higher (1=way too little; 3=just right; 5=way too much)

Rating	No. of respondents with this rating	No. of elements accommodated by repositories with this rating	Average no. of elements accommodated by repository with this rating
1 (way too little)	7	2-4	2.6
2	12 <sup>55</sup>	2-8	3.8
3 (just right)	15 <sup>56</sup>	1-10	3.9
4	4	3-11	7
5 (way too much)	0	n/a	n/a
Don't know	3	1-7	3.7

Table 6. Rating of technical metadata capability relative to number of technical metadata elements accommodated



<b>Preservation metadata you <i>can</i> collect is ...</b>						
No. of elements accommodated by the repository	1 “way too little”	2	3 “just right”	4	5 “way too much”	0 “don’t know”
1 (8 respondents)	3	3	2	0	0	0
2-3 (7 respondents)	0	3	1	0	1	2
4-11 (9 respondents)	0	2	3	3	0	1

Table 7. Ratings of preservation metadata capability

Number of elements	Number of respondents rating 3 or higher	Percent of respondents rating 3 or higher
1 (8 respondents)	2	25 percent
2-3 (7 respondents)	2	29 percent
4-11 (9 respondents)	6	66 percent

Table 8. Percent of respondents rating preservation metadata 3 ('just right') or higher (1=way too little; 3=just right; 5=way too much)

Rating	No. of respondents with this rating	No. of elements accommodated by repositories with this rating	Average no. of elements accommodated by repository with this rating
1 (way too little)	3	1	1
2	8	1-10	3
3 (just right)	6	1-11	4.2
4	3	5-10	7.3
5 (way too much)	1	2	2
Don't know	3	2-4	3

Table 9. Rating of preservation metadata capability relative to number of preservation metadata elements accommodated

Number of rights elements accommodated	Average number of Object Use tasks supported
0 (11 respondents)	.5
1-2 (13 respondents)	1.2
3-4 (13 respondents)	2.4
5-18 (16 respondents)	2.2

Table 10. Object Use tasks and rights metadata: average number of supported tasks relative to number of rights elements accommodated

Number of technical elements accommodated	Average number of Fixity/Integrity/Authenticity tasks supported
0 (12 respondents)	2.7
1-2 (12 respondents)	4
3 (12 respondents)	3.9
4 (8 respondents)	6.5
5-11 (9 respondents)	6.8

Table 11. Fixity/Integrity/Authenticity tasks and technical metadata: average number of supported tasks relative to number of technical metadata elements accommodated

Number of preservation elements accommodated	Average number of preservation tasks supported
0 (10 respondents)	1.4
1-3 (10 respondents)	0.8
4-6 (13 respondents)	2.5
7-8 (10 respondents)	4.7
9-24 (10 respondents)	6.4

Table 12. Preservation tasks and preservation metadata: average number of supported tasks relative to number of preservation metadata elements accommodated

## THE SURVEY

*[Note: An asterisk preceding a survey question indicates that a response was required.]*

### Introduction

You are invited to participate in a research study about administrative metadata. The purpose of this research is to identify the administrative metadata collected by ARL repositories and to determine how that metadata is currently used. This survey addresses administrative metadata of four types: rights metadata, technical metadata, source metadata, and preservation metadata.

A snapshot of current ARL practices identifying commonly collected data should help define core administrative metadata, stimulate the development of best practices and tools, as well as encourage discussion of collaborative uses of administrative metadata, to ensure longevity of ARL digital collections.

We are looking for a single institutional response. If you have more than one repository, please answer the questions for your primary repository, the one housing your organization's scholarship, and/or digital collections, i.e., the one most aligned with your institutional mission. If your organization hosts a consortial repository, please coordinate a single response from a representative of that repository. In some cases an organization will have one response for its own repository and a second response for a consortial repository.

Answers are saved as you respond. You can stop at any time and return to the survey later \*if you enable cookies in your browser.\*

This survey is being conducted by Rhonda J. Marker and Jane Johnson Otto at Rutgers University.

We thank you for participating!

### Consent Form

This research is confidential. The research records will include some information about you and this information will be stored in such a manner that some linkage between your identity and the response in the research exists. The information collected about you includes name, email, and position title (all optional), and organization name. Please note that we will keep this information confidential by limiting individual's access to the research data and keeping it in a secure location.

The research team and the Institutional Review Board (a committee that reviews research studies in order to protect research participants) at Rutgers University are the only parties that will be allowed to see the data, except as may be required by law. If a report of this study is published, or the results are presented at a professional conference, only group results will be stated. All study data will be kept for five years.

There are no foreseeable risks to participation in this study. In addition, you may receive no direct benefit from taking part in this study.

The survey will take about 15-20 minutes to complete, depending on responses. Participation in this study is voluntary. You may choose not to participate, and you may withdraw at any time during the study procedures without any penalty to you. In addition, you may choose not to answer any questions with which you are not comfortable.

If you have any questions about the study or study procedures, you may contact me at

Jane Otto  
Media and Music Metadata Librarian  
Rutgers University Libraries  
jjotto@rulmail.rutgers.edu  
(848) 445-5950  
(732) 445-5888 (fax)  
Rutgers, the State University of New Jersey  
Library Technical Services Building  
47 Davidson Road  
Piscataway, NJ 08854-5603

If you have any questions about your rights as a research subject, you may contact the IRB Administrator at Rutgers University at:

Rutgers University, the State University of New Jersey  
Institutional Review Board for the Protection of Human Subjects  
Office of Research and Sponsored Programs  
3 Rutgers Plaza  
New Brunswick, NJ 08901-8559  
Tel: 732-932-0150 ext. 2104  
Email: humansubjects@orsp.rutgers.edu

\*1. By completing this survey you agree to be a study subject. Please click "YES" to continue the survey. If you do not agree with the consent form and wish not to participate in this project, please click "No" to exit from this survey.

- Yes, I agree to participate
- No, I do not agree to participate

### **About Your Repository**

2. Please tell us the name of your primary repository, the one housing your organization's scholarship, and/or digital collections, i.e., the one most aligned with your institutional mission (for example, Deep Blue). All questions in the survey will pertain to the repository named here.

\*3. Please provide the URL for your repository's public homepage.

\*4. This repository is:

- "the" institutional repository (of broad scope but limited to the organization's scholarly output)
- the organization's digital library, limited to digitized library or archive collections
- a repository combining scholarly output with digitized library/archive collections
- an explicitly subject- or format-specific repository not falling into one of the above categories
- Other (please specify)

\*5. Which repository software does this repository use?

- Archimede
- contentDM
- CWIS
- Digital Commons
- DigiTool
- DSpace
- Eprints



- ETD-db
- Fedora
- Greenstone
- IR+
- Red Hat
- Other (please specify)

\*6. About how many fully cataloged resources do you add annually to your repository?

- Under 1000
- 1000-5000
- Over 5000

Comments:

\*7. What types of resources can this repository currently accept? Choose all that apply.

- Texts (e.g., books, letters, dissertations, periodicals)
- Still images (e.g., photographs, graphics, maps)
- Video
- Audio
- Data sets
- None of these
- Other (please specify)

### Your Administrative Metadata

\*8. If you incorporate elements from any of these metadata standards for your repository's rights, technical, source, or preservation metadata, please indicate which. Choose all that apply.

- Simple Dublin Core
- Qualified Dublin Core
- MODS
- MARC
- PREMIS
- NISO MIX (Z39.87)
- PBCore
- MPEG-7
- California Digital Library copyrightMD
- LC AV Prototype- text schema
- LC AV Prototype- AMD (audio) schema
- LC AV Prototype- VMD (video) schema
- LC AV Prototype- IMD (analog image) schema
- LC AV Prototype- RMD (rights) schema
- LC AV Prototype- PMD (digiprov) schema
- None of the above
- Other (please specify)

9. If the metadata your repository collects is documented and publicly available, what is the URL?

10. If metadata your repository collects is documented but not freely available on the Web, would you be willing to share the documentation with us? (Please send to [jjotto@rulmail.rutgers.edu](mailto:jjotto@rulmail.rutgers.edu).)

- yes
- no
- not applicable

Comments:

## Rights Metadata

\*11. Does your repository accommodate rights metadata? Rights metadata is information about intellectual property rights granted or reserved, copyright holder or licensor, etc.

- yes
- no (skip to technical metadata)
- I don't know

Comments:

\*12. Indicate which rights elements your metadata scheme \*accommodates\*

*Check the box if your metadata scheme has an element dedicated to that information, or a more granular form of it. If your metadata scheme has no \*dedicated\* element for that information, LEAVE THE BOX UNCHECKED, whether or not the scheme can accommodate the information elsewhere (for example, in a note). We are trying to determine 1) what data can be collected and 2) the granularity of repository metadata schemes.*

- Rights statement or license terms
- Rights granted the repository (replicate, migrate, modify, use, delete, etc.)
- Copyright status (e.g., copyright protected, public domain)
- Copyright jurisdiction
- Statute citation
- Statute jurisdiction
- Date of original copyright
- Date of copyright renewal
- Rights basis (copyright, license, statute)
- Copyright notice as it appears on the resource
- Availability status (e.g., open, restricted, unavailable)
- Rationale for availability status (e.g., deed of gift)
- Publication status (e.g., published, unpublished, unknown)
- Indication if watermarked
- Agent name (e.g., Rightsholder)
- Agent contact information (e.g., Rightsholder contact information)
- Note(s) about rights
- Link(s) to rights documentation
- None of the above
- Other (please specify)

\*13. For what percentage of your repository objects (roughly) do you record some rights metadata of the types mentioned above?

- 0%
- More than zero but fewer than one third of the objects
- Between one and two thirds of the objects
- More than two thirds of the objects, but not 100%
- 100% of the objects
- I don't know

Comments:

\*14. Given what your repository is meant to do

Rights metadata you CAN collect is

Way too little             Just right             Way too much       I don't know

Rights metadata you DO (routinely) collect is

- Way too little             Just right             Way too much       I don't know

## Technical Metadata

\*15. Does your repository accommodate technical metadata? Technical metadata describes the characteristics of the archival digital file, e.g., file size, compression scheme, operating system, codec, etc.

- yes
- no (skip to technical metadata)
- I don't know

Comments:

\*16. Indicate which technical metadata elements your metadata scheme \*accommodates\*

*Check the box if your metadata scheme has an element dedicated to that information, or a more granular form of it. If your metadata scheme has no \*dedicated\* element for that information, LEAVE THE BOX UNCHECKED, whether or not the scheme can accommodate the information elsewhere (for example, in a note). We are trying to determine 1) what data can be collected and 2) the granularity of repository metadata schemes.*

- Levels of encoding or encryption applied
- Fixity check data
- File size
- Format (.pdf, .htm)
- Creating application name [and/or version]
- Information on access inhibitors (encryption, password protection)
- Technical metadata notes
- Compression data (whether or not compressed, compression ratio etc.)
- Color space
- Capture information (who did it, scanner/camera details, etc.)
- Orientation (e.g., landscape or portrait, degrees of rotation)
- Bits per sample (8-bit; 16-bit, etc.)
- Embedded application data
- None of the above
- Other (please specify)

\*17. Indicate which ADDITIONAL technical metadata elements for VIDEO your scheme accommodates. These elements apply to the archival digital object itself.

*Check the box if your metadata scheme has an element dedicated to that information, or a more granular form of it. If your metadata scheme has no \*dedicated\* element for that information, LEAVE THE BOX UNCHECKED, whether or not the scheme can accommodate the information elsewhere (for example, in a note). We are trying to determine 1) what data can be collected and 2) the granularity of repository metadata schemes.*

- We don't accept video in our repository
- Time code
- Duration
- Signal format (NTSC, PAL, etc.)
- Codec information (name, version, creating app, etc.)
- Bit rate information (kBs, whether fixed or variable, etc.)
- Sampling information (sampling rate, bit depth, word size, etc.)
- Video encoding scheme
- Byte order (little endian or big endian)
- Frame information (height & width, aspect ratio, frame rate)
- Presence of sound

- Audio channel data (no. of channels, left-right position, etc.)
- Audio presentation (stereo, mono, etc.)
- Audio codec information (name, version, creating app, etc.)
- Audio bit rate information (kBps, whether fixed or variable, etc.)
- Audio sampling information (sampling rate, bit depth, word size, etc.)
- None of the above
- Other (please specify)

\*18. Indicate which ADDITIONAL technical metadata elements for AUDIO your scheme accommodates. These elements apply to the archival digital object itself.

*Check the box if your metadata scheme has an element dedicated to that information, or a more granular form of it. If your metadata scheme has no \*dedicated\* element for that information, LEAVE THE BOX UNCHECKED, whether or not the scheme can accommodate the information elsewhere (for example, in a note). We are trying to determine 1) what data can be collected and 2) the granularity of repository metadata schemes.*

- We don't accept audio in our repository
- Time code
- Duration
- Audio encoding (e.g., PCM)
- Byte order (little endian or big endian)
- First sample offset (number of bytes immediately prior to the first byte of audio data)
- Information on audio data blocks
- Audio channel data (no. of channels, left-right position, etc.)
- Audio presentation (stereo, mono, etc.)
- Audio codec information (name, version, creating app, etc.)
- Audio bit rate information (kBps, whether fixed or variable, etc.)
- Audio sampling information (sampling rate, bit depth, word size, etc.)
- None of the above
- Other (please specify)

\*19. Which of these metadata extraction tools do you employ?

- We do not use metadata extraction tools
- EMET (ARTstor)
- Exiftool
- File identifier (Optima SC Inc.)
- Jhove
- Metadata extraction tool (National Library of New Zealand)
- Other (please specify):

\*20. For what percentage of your repository objects (roughly) do you record some technical metadata of the types mentioned above?

- 0%
- More than zero but fewer than one third of the objects
- Between one and two thirds of the objects
- More than two thirds of the objects, but not 100%
- 100% of the objects
- I don't know

Comments:

\*21. Given what your repository is meant to do

Technical metadata you CAN collect is

Way too little       Just right       Way too much       I don't know

Technical metadata you DO (routinely) collect is

Way too little       Just right       Way too much       I don't know

## Preservation Metadata

\*22. Does your repository accommodate preservation metadata? Preservation metadata is that which supports the digital preservation process, beyond digital characteristics of the archival file.

- yes
- no (skip to source metadata)
- I don't know

Comments:

\*23. Indicate which preservation elements your scheme accommodates.

*Check the box if your metadata scheme has an element dedicated to that information, or a more granular form of it. If your metadata scheme has no \*dedicated\* element for that information, LEAVE THE BOX UNCHECKED, whether or not the scheme can accommodate the information elsewhere (for example, in a note). We are trying to determine 1) what data can be collected and 2) the granularity of repository metadata schemes.*

- Preservation level (bit-level, full, etc.)
- Significant properties (properties to be preserved, such as content, appearance, structure, behavior, context)
- Storage location (file location, location scheme (e.g., handle, URI), storage medium (hard disc, mag tape, etc.)
- Hardware/software supporting use of the object (including operating system and associated files required to use or render the object)
- Digital signature (signature itself, its encoding, encryption/hash algorithms, etc.)
- Preservation action (e.g., migration)
- Preservation outcome
- Person/organization responsible for preservation action
- Software associated with preservation action
- Generation or use type (preservation master, production master, etc.)
- Embedded application data
- Links between objects when one is derived from the other
- Condition evaluation
- None of the above
- Other (please specify)

\*24. For what percentage of your repository (roughly) do you record some preservation metadata of the types mentioned above?

- 0%
- More than zero but fewer than one third of the objects
- Between one and two thirds of the objects
- More than two thirds of the objects, but not 100%
- 100% of the objects
- I don't know

Comments:

- \*25. Given what your repository is meant to do
- Preservation metadata you CAN collect is
- Way too little             Just right             Way too much       I don't know
- Preservation metadata you DO (routinely) collect is
- Way too little             Just right             Way too much       I don't know

### Source Metadata

\*26. Does your repository accommodate source metadata? Source metadata documents the physical characteristics of the original (usually analog) PHYSICAL source object from which the digital master is derived (for example, an original film negative or vinyl record). It might include elements such as dimensions, sound and color characteristics, etc.

- yes
- no (skip source metadata questions)
- I don't know

Comments:

\*27. Which of the following do you record in your source metadata?

Physical characteristics of the object (extent, color characteristics, sound characteristics for audio, gauge for film and video, etc.)

- Provenance information
- Physical location
- None of the above
- Other (please specify)

\*28. Thinking only of those digital repository resources for which you also hold the physical source object, for what percentage of those source objects (roughly) do you record some source metadata?

*We are asking about elements for \*physical characteristics\* of the original source object; this question is NOT about descriptive metadata.*

- 0%
- More than zero but fewer than one third of the objects
- Between one and two thirds of the objects
- More than two thirds of the objects, but not 100%
- 100% of the objects
- I don't know

Comments:

\*29. Given what your repository is meant to do

- Source metadata you CAN collect is
- Way too little             Just right             Way too much       I don't know
- Source metadata you DO (routinely) collect is
- Way too little             Just right             Way too much       I don't know

## How Your Metadata is Used

\*30. Which of these tasks does your metadata support?

- Store original digital object
- Protect against data corruption & loss
- Ensure authenticity of digital resources over time
- Track origins and chain of custody of digital object (provenance)
- Document digital object's bitstream over the long term
- Ensure resources will survive and continue to be understandable into the long term
- Maintain vigorous and ongoing testing and validation program to ensure independent understandability of the digital object
- Enable format migration/transformation upon obsolescence
- Track migration path of digital object & any changes over time (digital provenance)
- Document owner of intellectual property rights
- Document permissions to distribute, duplicate, transfer, and/or alter (e.g., through migration)
- Store contractual agreements pertaining to rights and permissions
- Facilitate condition assessment of digital preservation master
- Facilitate decision-making for preservation managers
- Document preservation actions taken
- Document effects of preservation strategies
- Document relationships between multiple manifestations of a digital object
- Document details of original source object (e.g., provenance, preservation, condition)
- Enable repurposing of content (e.g., for revenue generation)
- None of the above
- Other (please specify)

\*31. Who creates your administrative metadata? (Check all that apply.)

- Depositor
- Cataloging and metadata department
- Digital repository department
- Special collections department
- Metadata extraction tool
- System supplied defaults
- None of the above
- Other (please specify)

\*32. Does your organization have staffing sufficient to perform the tasks you checked?

- yes
- no
- I don't know
- not applicable

Comments:

## Your Information

33. Please tell us your name.

34. Please tell us your email address.

\* 35. Your organization:

36. Your title:

\*37. If you are developing best practices that might be useful to explore further, we would like to follow up with you for more information, and will send you a preview of our findings. May we follow up with you for further information or clarification?

- yes
- no

**Thank you!**

Thank you for participating in our survey. We very much appreciate your contribution. If you have any questions, please feel free to contact us.

Jane Johnson Otto  
jjotto@rulmail.rutgers.edu

Rhonda J. Marker  
rmarker@rulmail.rutgers.edu



## References and Notes

---

<sup>1</sup> Metadata “is widely acknowledged to be crucial to the long-term preservation of digital entities” (*Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation*, 2003: viii, accessed July 14, 2013,

[http://eprints.erpanet.org/94/01/NSF\\_Delos\\_WG\\_Pres\\_final.pdf](http://eprints.erpanet.org/94/01/NSF_Delos_WG_Pres_final.pdf)); “digital preservation metadata is the information that is essential to ensure long-term accessibility of digital resources” (Angela Dappert and Adam Farquhar, “Implementing Metadata that Guides Digital Preservation Services,” in *iPRES 2009: the Sixth International Conference on Preservation of Digital Objects* (California Digital Library, Office of the President, 2009): 50, accessed July 14, 2013, <http://escholarship.org/uc/item/12p437bd>; “reliable authentic digital objects will not be preserved across time without adequate preservation metadata” (Wendy Duff, “Metadata in Digital Preservation: Foundations, Functions and Issues,” in M. Bischoff, H. Hofman, and S. Ross, *Metadata in preservation: Selected papers from ERPANET seminar at the Archives School Marburg, 3-5 September 2003* (Veröffentlichungen der Archivschule Marburg, Institute für Archivwissenschaft, Nr. 40): 27, cited in Steve Knight, “Preservation Metadata: National Library of New Zealand Experience.” *Library Trends* 54, no. 1 (Summer 2005): 96. A good overview of preservation metadata, including what that term encompasses, may be found in Brian Lavoie and Richard Gartner, *Preservation Metadata: DPC Technology Watch Report no. 05-01: September 2005*, accessed July 14, 2013, [www.dpconline.org/docs/reports/dpctw05-01.pdf](http://www.dpconline.org/docs/reports/dpctw05-01.pdf).

<sup>2</sup> *Feasibility and Requirements Study on Preservation of E-Prints: Report Commissioned by the Joint Information Systems Committee (JISC)* (October 29, 2003): 5, accessed July 14, 2013, [www.jisc.ac.uk/media/documents/programmes/preservation/e-prints\\_report\\_final.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/e-prints_report_final.pdf).

---

<sup>3</sup> *Trustworthy Repositories Audit & Certification: Criteria and Checklist* (Chicago, Illinois: Center for Research Libraries, February 2007): 28, accessed July 14, 2013, [www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf).

<sup>4</sup> *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA: RLG, May 2002: 25, accessed July 14, 2013, [www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf](http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf).

<sup>5</sup> PREMIS' core elements are applicable to objects in all formats, but lack the detailed, format-specific technical metadata that is "clearly necessary for implementing most preservation strategies." *PREMIS Data Dictionary for Preservation Metadata*, version 2.0 (March 2008): 24, accessed July 14, 2013, [www.loc.gov/standards/premis/v2/premis-2-0.pdf](http://www.loc.gov/standards/premis/v2/premis-2-0.pdf).

<sup>6</sup> For example, MPEG-7: ISO/IEC 15938, *Multimedia Content Description Interface*, accessed July 14, 2013, <http://mpeg.chiariglione.org/standards/mpeg-7>.

<sup>7</sup> For example, AES57-2011: *AES Standard for Audio Metadata: Audio Object Structures for Preservation and Restoration*. New York: Audio Engineering Society, 2011.

<sup>8</sup> For example, *PREMIS Data Dictionary*, 2008.

<sup>9</sup> Most obviously, Dublin Core: *DCMI Metadata Terms*, 2012, accessed July 14, 2013, [dublincore.org/documents/dcmi-terms](http://dublincore.org/documents/dcmi-terms).

---

<sup>10</sup> Chen, Mingyu and Michele Reilly. "Implementing METS, MIX, and DC for Sustaining Digital Preservation at the University of Houston Libraries." *Journal of Library Metadata* 11, no. 2 (Apr, 2011): 95, doi: [dx.doi.org/10.1080/19386389.2011.570662](https://doi.org/10.1080/19386389.2011.570662).

<sup>11</sup> *Trusted Digital Repositories: Attributes and Responsibilities*, May 2002.

<sup>12</sup> Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *Portal: Libraries and the Academy* 3, no. 2 (April 2003): 327-336.

<sup>13</sup> "Feasibility and Requirements Study on Preservation of E-Prints: Report Commissioned by the Joint Information Systems Committee (JISC)." (October 29, 2003): accessed July 14, 2013, [www.jisc.ac.uk/media/documents/programmes/preservation/e-prints\\_report\\_final.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/e-prints_report_final.pdf).

<sup>14</sup> nector: Network of Expertise in long-term STORAge Working Group on Trusted Repositories Certification. Catalogue of Criteria for Trusted Digital Repositories. Frankfurt am Main: Deutsche Nationalbibliothek, June 2006.

<sup>15</sup> *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, February 2007.

<sup>16</sup> For example, Charles W. Bailey et al. found that preservation is viewed as one of the top three benefits of the institutional repository (*SPEC Kit 292: Institutional repositories* (Washington, D.C.: Association of Research Libraries, 2006): 21, accessed July 14, 2013, <http://publications.arl.org/Institutional-Repositories-SPEC-Kit-292/21>). Library leaders have not only ranked preservation high, but they perceive preservation issues as a key benefit for potential repository users and contributors. See Elizabeth Yakel et

---

al., "Institutional Repositories and the Institutional Repository: College and University Archives and Special Collections in an Era of Change." *The American Archivist*, 71 (Fall/Winter 2008): 339-40.

<sup>17</sup> *Invest to Save*, 2003: 2.

<sup>18</sup> Yakel et al., "Institutional Repositories and the Institutional Repository," 2008: 339.

<sup>19</sup> Seamus Ross, "Challenges to Digital Preservation and Building Digital Libraries," *World Library and Information Congress: 69<sup>th</sup> IFLA General Conference and Council, 1-9 August, 2003, Berlin*: accessed July 14, 2013, [http://eprints.erpanet.org/104/01/ROSS\\_IFLABERLIN2003\\_209e-Ross.pdf](http://eprints.erpanet.org/104/01/ROSS_IFLABERLIN2003_209e-Ross.pdf): 6-7.

<sup>20</sup> Steve Knight, "Preservation Metadata: National Library of New Zealand Experience." *Library Trends* 54, no. 1 (Summer 2005): 95.

<sup>21</sup> Kathleen Shearer, "The CARL Institutional Repositories Project: A Collaborative Approach to Addressing the Challenges of IRs in Canada," *Library Hi Tech* 24, no. 2 (2006): 170.

<sup>22</sup> Mary Westell, "Institutional Repositories: Proposed Indicators of Success." *Library Hi Tech* 24, no. 2 (2006): 222.

<sup>23</sup> DigitalPreservationEurope. "DPE Research Roadmap, DPE-D7.2." (June 2007), accessed July 14, 2013, [www.digitalpreservationeurope.eu/publications/dpe\\_research\\_roadmap\\_D72.pdf](http://www.digitalpreservationeurope.eu/publications/dpe_research_roadmap_D72.pdf): 28, 19, 20.

---

<sup>24</sup> For example, the JISC report, in citing the high cost of metadata, notes that “it is notoriously difficult to predict” preservation costs, in part due to the lack of practical experience on which to base cost estimates (*Feasibility and Requirements Study*, 2003): 41.

<sup>25</sup> See, for example, *Trusted Digital Repositories: Attributes and Responsibilities*, May 2002: 19-20, and Raym Crow, *The Case for Institutional Repositories: A SPARC Position Paper*. Washington, D.C.: SPARC, 2002, accessed July 14, 2013, [www.sparc.arl.org/bm~doc/ir\\_final\\_release\\_102.pdf](http://www.sparc.arl.org/bm~doc/ir_final_release_102.pdf): 27-28.

<sup>26</sup> *Feasibility and Requirements Study*, 2003: 5. The blueprint comment appears on page 35.

<sup>27</sup> Knight, "Preservation Metadata: National Library of New Zealand Experience," 2005: 97.

<sup>28</sup> D. Alemneh, “Barriers to adopting PREMIS in cultural heritage institutions: an exploratory study,” *Archiving 2008 Final Program and Proceedings, Washington, DC*: 71-80, cited in Devan Ray Donaldson and Paul Conway, “Implementing PREMIS: a case study of the Florida Digital Archive,” *Library Hi Tech*, 28, issue 2 (2010): 276.

<sup>29</sup> Caplan, Priscilla. "Preservation Metadata," in Seamus Ross and M. Day, *DCC Digital Curation Manual*, July 2006: 23, 26.

<sup>30</sup> Angela Dappert and Markus Enders. "Digital Preservation Metadata Standards," *Information Standards Quarterly* 22, no. 2 (Spring 2010): 5-13.

<sup>31</sup> Dappert and Farquhar. *Implementing Metadata that Guides Digital Preservation Services*, 2009: 50.

---

<sup>32</sup> "Preservation Metadata," *Paradigm*, accessed July 14, 2013,

[www.paradigm.ac.uk/workbook/metadata/preservation-considerations.html](http://www.paradigm.ac.uk/workbook/metadata/preservation-considerations.html).

<sup>33</sup> See, for example, LDP Centre. *CODA-META: Curation of Digital Assets-Metadata* (Boden, Sweden: LDB-cemtri, 2008) and Bronwyn Lee, Gerard Clifton, and Somaya Langley, *Australian Partnership for Sustainable Repositories PREMIS Requirement Statement Project Report* (National Library of Australia, July 2006): 28-31.

<sup>34</sup> For studies of descriptive metadata or metadata more generally, see, for example, Eun G. Park and Marc Richard, "Metadata Assessment in e-Theses and Dissertations of Canadian Institutional Repositories," *The Electronic Library* 29, no. 3 (2011): 394-407; Jung-ran Park and Yuji Tosaka, "Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability," *Information Technology and Libraries* 29, no. 3 (2010): 104-116; and Jin Ma, "Metadata in ARL Libraries: A Survey of Metadata Practices." *Journal of Library Metadata* 9, no. 1-2 (2009): 1-14. Lopatin surveyed library metadata practices for digital projects, but was not necessarily surveying repositories (see Laurie Lopatin, "Metadata Practices in Academic and Non-Academic Libraries for Digital Projects: A Survey," *Cataloging & Classification Quarterly* 48, no. 8 (2010): 716-742).

<sup>35</sup> Gail McMillan, Matt Schultz, and Katherine Skinner, *SPEC Kit 325: Digital Preservation* (Washington, D.C.: Association of Research Libraries, October 2011). Findings regarding metadata schemas in use reflect the findings of this paper; see p. 11.

<sup>36</sup> ROAR and OpenDOAR both attempt to provide up-to-date and comprehensive listings of open access repositories. ROAR is hosted by the School of Electronics and Computer Science at the University of

---

Southampton; it lists repositories worldwide, and provides information on the growth and status of repositories in an effort to promote the development of open access. OpenDOAR is a directory of academic open access repositories, run by the Centre for Research Communications (CRC).

<sup>37</sup> In the questionnaire, “digital library” was defined as a repository “limited to digitized library or archive collections,” but in hindsight should have included born-digital library collections. Therefore responses fitting the broader definition were included in the “digital library” category.

<sup>38</sup> Several respondents noted caveats. One organization’s repository could accept media files “but lacks dynamism to stream”; files can simply be uploaded to the repository and downloaded by users. Two organizations noted that while their repositories could accept many formats, the extent to which those files are preserved is format-dependent. Another noted that a set of files might be archived, but not necessarily as a “supported live object that can execute scripts.” Of course repositories can *accommodate* metadata for many types of materials, but this does not mean the repositories actually *house* all those types of materials. The survey did not ask for number of the various types of resources actually housed.

<sup>39</sup> When this survey was originally drafted, AES57-2011 was still in draft form. Due to an oversight, it was never added into the list of schema options before the survey was distributed to respondents. For the record, AES57-2011 was listed under “Other” by one repository.

<sup>40</sup> One respondent selected “none of the above” and declined to specify what standards were incorporated into the repository schema; this respondent was excluded from the count of total respondents to this question.

---

<sup>41</sup> Of fifty-four total respondents, fifty-three named their metadata schema; one selected “none of the above” and did not specify.

<sup>42</sup> All other standards were incorporated by fewer than three organizations, and included the Text Encoding Initiative (TEI), Visual Resources Association (VRA) Core, Library of Congress AV Prototype for Audio, Library of Congress AV Prototype for Video, PBCore. A number were incorporated by just one organization: AES57, DocumentMD, FITS, harvestMD, hulDrsAdmin, hulDrsRights, textMD, MPEG-7, California Digital Library, copyrightMD, Library of Congress AV Prototype for text, Analyzed Layout and Text Object (ALTO), FOXML (for capturing audit history), and Darwin Core.

<sup>43</sup> One respondent answered “I don’t know.” On the basis of this respondent’s answers to other rights metadata questions, the response was converted to “no” and the respondent’s answers to rights metadata questions 12-14 were accordingly deleted from the data set.

<sup>44</sup> Based on two comments associated with the rights metadata questions, it is likely that batchloaded legacy metadata may account, in part, for the dearth of rights metadata. Two additional respondents’ comments belied a perception that rights metadata is not necessary if all resources are public domain, or otherwise of a kind. The inclusion of consortial repositories in the survey further complicates data analysis here, since it is not always possible to dictate policy and practice to individual consortium members.

<sup>45</sup> Of forty-three repositories accommodating technical metadata, two did not specify elements and were therefore deleted from this data set.



---

<sup>46</sup> There were thirty-five valid responses to this question. Four respondents replied “I don’t know” to one or both parts of the question and three respondents appeared to misinterpret the question; these seven respondents were deleted from this data set. (Although what *can be* collected was meant to refer to system capability, three respondents rated what actually *is* collected higher than what *can be* collected, suggesting an alternative interpretation of the question.)

<sup>47</sup> Fifteen respondents rated the repository capability “just right” *and* specified the number of technical metadata elements accommodated. One respondent failed to specify and that response was deleted from this data set.

<sup>48</sup> Of the thirty respondents who said their repositories accommodate some preservation metadata, three qualified their “yes” with comments which roughly equated to “coming soon.” One such respondent was able to specify actual elements in Question 23, so that “yes” answer was retained even though it appeared that full implementation had not yet occurred; in the other two cases, no elements were specified in Question 23, so the “yes” answer was converted to a “no.” After these adjustments, it can be said that twenty-eight of fifty-four surveyed repositories (52 percent) accommodate preservation metadata.

<sup>49</sup> Of the twenty-eight repositories accommodating preservation metadata, four declined to specify which elements were used, selecting either “none of the above” or “other,” without enumerating alternatives.

<sup>50</sup> Of the twenty-nine repositories recording source metadata, one respondent said “none of the above” but did not elaborate. Another, answering for a consortium, was unable to provide a clear picture of contributed source metadata and so did not specify information types. These two responses were deleted from the data set.

---

<sup>51</sup> There were twenty-five valid responses to the question. Two respondents replied “I don’t know” and one misinterpreted the question, basing the percentage on a mix of born digital and digitized content; these responses were removed from the data.

<sup>52</sup> The list of tasks is a synthesis of information from a number of sources, including *PREMIS Data Dictionary*, March 2008; *Trustworthy Repositories Audit & Certification*, 2007; OCLC/RLG Working Group on Preservation Metadata, *Preservation Metadata for Digital Objects: a Review of the State of the Art*, 2001, accessed July 14, 2013, [www.oclc.org/content/dam/research/activities/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/content/dam/research/activities/pmwg/presmeta_wp.pdf); *Reference Model for an Open Archival information System (OAIS) : Recommended Practice* (Washington, D.C.: Consultative Committee for Space Data Systems, 2012), accessed July 14, 2013, <http://public.ccsds.org/publications/archive/650x0m2.pdf>; *Feasibility and Requirements Study*, 2003.

<sup>53</sup> Fifty-three of the fifty-four total respondents specified tasks supported by their metadata. (One respondent chose “None of the above” but specified no alternative and this response was pulled from the data set in calculating percentages and averages for this question.)

<sup>54</sup> For example, Hockx-Yu cites the “need for new, shared preservation services and information infrastructure” and describes the Joint Information Systems Committee’s vision of shared preservation services (Helen Hockx-Yu, “Digital Preservation in the Context of Institutional Repositories,” *Program: Electronic Library and Information Systems* 40, no. 3 (2006): 237, 239-241); Dappert and Enders describe digital content preservation as a “collaborative effort,” mentioning content sharing and a specific implementation to allow exchange of complex objects between heterogeneous preservation systems (TIPR, Towards Interoperable Preservation Repositories) (Angela Dappert and Markus Enders. "Digital Preservation Metadata Standards," *Information Standards Quarterly* 22, no. 2 (Spring 2010): 12-13);

---

DigitalPreservationEurope's *Research Roadmap* recommends the digital preservation community focus on developing services which "support the work of collaborative and distributed archival and preservation teams" (DigitalPreservationEurope. "DPE Research Roadmap, DPE-D7.2." (June 2007), accessed July 14, 2013, [www.digitalpreservationeurope.eu/publications/dpe\\_research\\_roadmap\\_D72.pdf](http://www.digitalpreservationeurope.eu/publications/dpe_research_roadmap_D72.pdf): 31).

<sup>55</sup> One of the 13 who assigned this rating failed to specify the metadata elements accommodated by the repository, so that respondent's response was omitted from this data.

<sup>56</sup> One of the 16 who assigned this rating failed to specify the metadata elements accommodated by the repository, so that respondent's response was omitted from this data.