

text content discoverable in and across information packages based on full-text as well as metadata-based queries using a powerful search server. The used repository provides instant access at the granularity of single files which can be viewed and/or packaged for dissemination using the provided E-ARK Web access components. The paper reports also on future directions to further improve the search capabilities of the system by employing data mining algorithms.

## 10. ACKNOWLEDGMENTS

Work presented in this paper is primarily supported by European Community's Seventh Framework Program through the E-ARK project under grant agreement number 620998.

## 11. REFERENCES

- [1] V. Borkar, M. J. Carey, and C. Li. Inside "big data management": Ogres, onions, or parfaits? In *Proceedings of the 15th International Conference on Extending Database Technology*, EDBT '12, pages 3–14, New York, NY, USA, 2012. ACM.
- [2] P. Caplan and R. S. Guenther. Practical preservation: the premis experience. *Library Trends*, 54(1):111–124, 2006.
- [3] CCSDS. *Reference Model for an Open Archival Information System (OAIS)*. CCSDS - Consultative Committee for Space Data Systems, January 2012. Version 2 of the OAIS which was published in June 2012 by CCSDS as "magenta book" (ISO 14721:2012).
- [4] K. Clemens. Geocoding with openstreetmap data. *GEOProcessing 2015*, page 10, 2015.
- [5] J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. In *OSDI'04: Proceedings of the 6th conference on operating systems design and implementation*. USENIX Association, 2004.
- [6] L. Faria, M. Ferreira, R. Castro, F. Barbedo, C. Henriques, L. Corujo, and J. C. Ramalho. Roda - a service-oriented repository to preserve authentic digital objects. In *Proceedings of the 4th International Conference on Open Repositories*. Georgia Institute of Technology, 2009.
- [7] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *In ACL*, pages 363–370, 2005.
- [8] S. Garfinkel, P. Farrell, V. Roussev, and G. Dinolt. Bringing science to digital forensics with standardized forensic corpora. *digital investigation*, 6:S2–S11, 2009.
- [9] A. J. Hey, S. Tansley, K. M. Tolle, et al. *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA, 2009.
- [10] A. J. Hey and A. E. Trefethen. The data deluge: An e-science perspective. 2003.
- [11] B. A. Jurik, A. A. Blekinge, R. B. Ferneke-Nielsen, and P. Møldrup-Dalum. Bridging the gap between real world repositories and scalable preservation environments. *International Journal on Digital Libraries*, 16(3):267–282, 2015.
- [12] R. King, R. Schmidt, C. Becker, and S. Schlarb. Scape: Big data meets digital preservation. *ERCIM News*, 89:30–31, 2012.
- [13] J. Lin, M. Gholami, and J. Rao. Infrastructure for supporting exploration and discovery in web archives.

In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pages 851–856, New York, NY, USA, 2014. ACM.

- [14] L. Medjkoune, S. Barton, F. Carpentier, J. Masanès, and R. Pop. Building scalable web archives. In *Archiving Conference, Archiving 2014 Final Program and Proceedings*, number 1, pages 138–143, 2014.
- [15] M. Noll. Benchmarking and stress testing an hadoop cluster with terasort, testdfsio & co. <http://www.michael-noll.com/blog/2011/04/09/benchmarking-and-stress-testing-an-hadoop-cluster-with-terasort-testdfsio-nbench-mrbench>, 4 2011.
- [16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [17] M. Radtisch, P. May, A. A. Blekinge, and P. Moldrup-Dalum. D9.1 characterisation technology, release 1 and release report. [http://www.scape-project.eu/wp-content/uploads/2012/03/SCAPE\\_D9.1\\_SB\\_v1.0.pdf](http://www.scape-project.eu/wp-content/uploads/2012/03/SCAPE_D9.1_SB_v1.0.pdf), 04 2012.
- [18] S. Schlarb, P. Cliff, P. May, W. Palmer, M. Hahn, R. Huber-Moerk, A. Schindler, R. Schmidt, and J. van der Knijff. Quality assured image file format migration in large digital object repositories. In J. Borbinha, M. Nelson, and S. Knight, editors, *iPRES 2013: 10th International Conference on Preservation of Digital Objects*, 2–6 September 2013, pages 9–16, Lisbon, Portugal, September 2013. Lisbon Technical University (IST).
- [19] R. Schmidt. An architectural overview of the scape preservation platform. In *9th International Conference on Preservation of Digital Objects*, pages 85–55. Citeseer, 2012.
- [20] R. Schmidt, M. Rella, and S. Schlarb. Constructing scalable data-flows on hadoop with legacy components. In *11th IEEE International Conference on e-Science, e-Science 2015, Munich, Germany, August 31 - September 4, 2015*, page 283, 2015.
- [21] R. Simon, E. Barker, L. Isaksen, and P. de Soto Cañamares. Linking early geospatial documents, one place at a time: Annotation of geographic documents with recogito. *e-Perimtron*, 10(2):49–59, 2015.

# Securing Trustworthy Digital Repositories

Devan Ray Donaldson  
School of Informatics and Computing School of Informatics and Computing  
Indiana University  
1320 E. 10<sup>th</sup> St., Wells Library 019  
Bloomington, IN 47405-3907  
+1-812-855-9723  
drdonald@indiana.edu

Raquel Hill  
School of Informatics and Computing  
Indiana University  
230E Lindley Hall  
Bloomington, IN 47406  
+1-812-856-5807  
ralhill@indiana.edu

Heidi Dowding  
Library Technologies  
Indiana University Libraries  
1320 E. 10<sup>th</sup> St., Wells Library W501  
Bloomington, IN 47405-3907  
+1-812-856-5295  
heidowdi@indiana.edu

Christian Keitel  
Landesarchiv Baden-Württemberg  
Eugenstraße 7  
D-70182 Stuttgart  
+49 0711-212-4276  
christian.keitel@la-bw.de

## ABSTRACT

Security is critical to repository trustworthiness. Recent international standards for Trustworthy Digital Repositories (TDRs) all specify some sort of security criteria that are necessary to adhere to in order to attain TDR status. However, little is known about how those who are responsible for addressing these criteria actually regard the concept of security. This study centers on digital repository staff members' perceptions of security, including their perceptions of security criteria in standards for TDRs. This paper discusses findings from surveys and semi-structured interviews with staff from repositories that have recently acquired the nestor seal of approval. We found that participants considered the principles of confidentiality, integrity, and availability as relevant to their notions of security. We also found that participants considered the security criteria required to acquire the nestor seal of approval as both sufficient and appropriate for addressing their repositories' needs. Implications for better understanding the security of digital repositories are discussed as well as directions for future research.

## Keywords

Security; Trustworthy Digital Repositories; Repository Staff Perceptions.

## 1. INTRODUCTION

Unarguably, security is part of what is necessary for a digital repository to be trustworthy. Evidence of the importance of security can be seen by examining criteria pertaining to security in recent standards for Trustworthy Digital Repositories (TDRs). For example, these criteria specify that staff identify sections of their repositories that are worthy of protection, analyze potential threats and perform risk assessment [4, 5, 9]. While security criteria in standards for TDRs seem relatively straightforward, little is known about actual staff members' perceptions of these security criteria. For example, staff may consider the criteria relatively easy to address, or they may consider the criteria rather challenging to address. Staff also may consider their repositories more secure as a result of adhering to these criteria or they may not. Digital repository staff members have a direct impact on the security of TDRs. They make decisions and implement policies that can result

either in increased security or compromises to security. For these reasons it is critically important to better understand how digital repository staff members think about security.

The purpose of this study is to understand digital repository staff members' perceptions of security for TDRs. The remainder of this paper is as follows. First, we explore scholarship on security in the digital preservation and computer science literatures. Second, the methodology section describes the sample of participants and explains why they were selected. The methodology section also describes data collection and analysis techniques. Third, the findings are reported. The paper concludes with a discussion and explication of implications of the study and recommends directions for future research.

## 2. SCHOLARSHIP ON SECURITY

### 2.1 Security in the Digital Preservation

#### Literature

Security refers to "the practice of defending information from unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording or destruction" [16, p. 224]. The best place to understand the phenomenon of security within the field of digital preservation is to examine recent standards for TDRs. They represent a consensus among key members of the digital preservation community on what constitutes best practice. They include specific criteria pertaining to security as part of attaining formal "trustworthy" status for digital repositories. For example, criterion C34 in DIN 31644 requires organizations and their infrastructures to protect their digital repositories and their contents [4, 11]. In particular, criterion C34 requires staff at organizations to protect the integrity of digital repositories and their content. To accomplish this, nestor certification against criterion C34 recommends that staff identify sections of the archive that are worthy of protection, analyze potential threats to the archive, and perform risk assessment "of the damage scenarios [to] ultimately result in a consistent security system" [11, p. 40]. For example, according to the explanatory notes on the nestor seal for TDRs, criterion C34 asks staff at organizations to identify which of three types of damage scenarios they perceive as a particular threat to information preserved by digital repositories: 1) malicious actions, 2) human error, or 3) technical failure. The explanatory notes also ask staff to consider the likelihood of each damage scenario, the seriousness of each scenario as well as what level

of residual risk is acceptable. Furthermore, they ask staff about what measures they are taking to counter these risks as well as how they plan to implement their risk analysis and planned countermeasures into their security systems. Finally, these notes ask staff about their plans to test and further develop their security systems.

Similarly to DIN 31644, ISO 16363 includes a section on security entitled “Security Risk Management” [9]. This section outlines security criteria for TDRs. According to ISO 16363, staff seeking “trustworthy” status for their digital repositories must maintain “a systematic analysis of security risk factors associated with data, systems, personnel, and physical plant” [9, p. 76]. A TDR must also:

- Implement controls to address defined security risks,
- Have delineated roles, responsibilities, and authorizations related to implementing changes within the system, and
- Have suitable written disaster preparedness and recovery plans.

ISO 16363 also describes three additional security concerns that could arise during audit. First, the auditor could be a false auditor or have malicious intent. Second, confidential information could be lost as a result of performing the audit, which could compromise the system. Third, damage to the repository system could occur while transferring information during audit. To guard against these security threats, recommendations in ISO 16363 include:

- Relying on repositories’ identification and authorization systems,
- Relying on the security systems of auditors and settling on information transfer agreements between repositories and auditors, and
- Relying on repositories’ security and safety systems.

Both DIN 31644 and ISO 16363’s security requirements draw upon an earlier standard for TDRs: Digital Repository Audit Method Based on Risk Assessment known as DRAMBORA [5]. For example, ISO 16363 recommends that digital repository staff members use DRAMBORA as a tool for performing risk assessments. Similarly, DIN 31644 recommends that digital repository staff members use DRAMBORA to help identify the sections of the archive which are worthy of protection, analyze any potential threats to the specific archive, and perform risk assessments of possible damage scenarios.

The DRAMBORA methodology consists of six steps. First, digital repository staff members should identify their objectives. DRAMBORA includes a list of examples of objectives for digital repository staff members to choose from. Second, digital repository staff members should identify the activities that are necessary to achieve their objectives and assets, including human resources and technological solutions, that are central to achieving repositories’ objectives. Third, digital repository staff members should align risks to their activities and assets. This step requires digital repository staff members to document the specific risks associated with each identified activity and asset. Here a single risk may associate with multiple activities, or vice versa. Fourth, digital repository staff members should assess, avoid, and treat risks by characterizing each risk’s “probability,

impact, owner, and the mechanisms or proposed mechanisms by which it can be avoided or treated” [5, p. 39]. Fifth, digital repository staff members should self-audit their repositories to determine what threats are most likely to occur and identify areas where improvement is required. Sixth, digital repository staff members should complete a risk register listing all identified risks and the results of their analysis and evaluation. Also known as a risk log, it should include information about the status of each risk and include details that can aid digital repository staff members in tracking and monitoring risks.

Taken together, standards for TDRs underscore the importance of security and provide relatively similar recommendations to digital repository staff members about how to address security. However, the security criteria themselves do nothing to illuminate actual digital repository staff members’ perspectives on security or their perceptions of the said security criteria.

## 2.2 Security in the Computer Science Literature

Relevant to a discussion on security in the digital preservation literature is discussion of security in the computer science literature. In digital preservation, the primary focus is on the security of digital repositories and their content. On the other hand, in the field of computer science security is more encompassing, including a broad range of computing infrastructures, not just digital repositories. Computer science also has a longer, more established body of literature on security, including definitions and metrics for the concept.

Computer scientists who specialize in security research have reached a consensus that computer security consists of at least three main principles: confidentiality, integrity, and availability. Confidentiality refers to concealment of information or resources, integrity refers to the trustworthiness of data or resources, and availability refers to the ability to use the information or resource desired [1]. While security researchers seem to agree on these three principles of security, others have proposed additional security elements. For example, some researchers have recommended including the concept of accountability, “the security goal that generates the requirement for actions of an entity to be traced uniquely to that entity,” in defining trustworthiness [17, p. A-1]. As another example, OECD guidelines proposed nine security principles: awareness, responsibility, response, ethics, democracy, risk assessment, security design and implementation, security management, and reassessment [12]. Stoneburner, Hayden, and Feringa [17] proposed thirty-three principles related to having a security foundation, risk, ease of use, increasing resilience, reducing vulnerabilities, and designing with the network in mind. Parker [13] extended the classic Confidentiality-Integrity-Availability (CIA) triad by adding three elements: possession, authenticity, and utility. After a thorough review of the literature, Cherdantseva and Hilton [3] proposed extending the CIA triad to an Information Assurance and Security (IAS) octave consisting of: confidentiality, accountability, auditability, authenticity/trustworthiness, non-repudiation, and privacy. It is important to note that Cherdantseva and Hilton had IAS academics and experts evaluate the IAS octave. According to Cherdantseva and Hilton, the IAS octave is part of a larger, all encompassing reference model of information assurance and security. Although alternative models of security exist, all seem to incorporate confidentiality, integrity, and availability at their core.

In addition to multiple definitions of security, the literature on security in computer science also offers some security metrics. For example, these metrics can provide assessment of security properties, measurement of adherence to secure coding standards, monitoring and reporting of security status, and gauge the effectiveness of various security controls [7, 10]. Although some security metrics exist, researchers acknowledge that security is actually quite difficult to measure. Pfleeger and Cunningham [14] list nine reasons why security is hard to measure:

- We can’t test all security requirements,
- Environment, abstraction, and context affect security,
- Measurement and security interact,
- No system stands alone,
- Security is multidimensional, emergent and irreducible,
- The adversary changes the environment,
- Measurement is both an expectation and an organizational objective,
- We’re overoptimistic, and
- We perceive gain differently from loss.

Common to both computer security and security for digital repositories is threat modeling. During the threat modeling process, assets are identified; threats against the assets are enumerated; the likelihood and damage of threats are quantified; and mechanisms for mitigating threats are proposed [2, 5, 8, 9, 11, 15].

While some components of the threat modeling process are qualitative, quantifying the risk of threats enables system administrators to rank the order in which threats should be addressed. Within the computer science literature, various approaches have been proposed for characterizing and quantifying the risk of threats, including calculating risk as the product of the damage potential and the likelihood of occurrence,  $Risk = Criticality * Likelihood of Occurrence$  [8]. Dread, an approach proposed by Microsoft, calculates risk across several categories, including: Damage potential, Reproducibility, Exploitability, Affected users, and Discoverability [8]. Using Dread, a threat is rated on a scale from 1 to 10 for each category, with the resulting risk being the average of all ratings. Butler and Fischbeck [2] propose a multiple attribute threat index (TI) for assessing the risk of a threat. TI captures the relative importance of each type of threat [2], where  $TI_a = Freq_a * (\sum_{j=attributes} W_j * X_{aj})$ , and  $W_j$  is the attribute weight and  $X_{aj}$  is the most likely outcome value for the threat.

While quantifying risks will enable us to capture an organization’s security requirements, Pfleeger [15] advises that we should avoid false precision by doing the following:

- Base the probability distribution of a threat/attack occurring on historical data, not just on expert judgment;
- Since “both scientists and lay people may underestimate the error and unreliability in small samples of data, particularly when the results are

consistent with preconceived, emotion-based beliefs”, we are to be mindful of the size of our experiments and the scalability of our results.

While measuring security is difficult, and few security metrics of any kind exist, metrics for understanding perceptions of security are particularly scant.

Taken together, the literature on security in digital preservation and computer science stress the importance of security, while also leaving several open research questions. This study focuses on four of them:

1. How do digital repository staff members think about the security of Trustworthy Digital Repositories?
2. What are digital repository staff members’ attitudes toward security criteria in standards for Trustworthy Digital Repositories?
3. How relevant are security principles that have been established in the computer science domain to digital repository staff members’ concept of security?
4. Is it possible to develop a survey that could serve as a tool for measuring digital repository staff members’ perceptions of security for Trustworthy Digital Repositories?

## 3. METHODS

To address the research questions, we conducted interviews with digital repository staff members at organizations whose repositories have attained formal, third-party trustworthy status. We also administered surveys to those individuals. The purpose of using these data collection methods was to understand how the participants thought about security and to assess measurement of the concept. While various standards for trustworthy digital repositories exist [4, 6, 9], at present, DIN 31644 is the only standard that: 1) has been formally recognized by a standards-granting body, and 2) has organizations whose repositories have been formally certified by third parties. Thus, we decided to include in our study only digital repository staff members whose repositories have recently acquired nestor seals of approval, signifying formal, third-party certification by the DIN 31644 standard. To date, two organizations have successfully acquired nestor seals of approval. During April 2016, we recruited participants at these institutions via email, asking them to participate in our interviews and take our survey.

### 3.1 Interviews

During semi-structured interviews, participants discussed their definitions of security and trustworthiness. They also discussed their views on the relationship between the trustworthiness of digital repositories and their security. Afterwards, participants discussed security criteria in DIN 31644 (e.g., criterion C34), including how easy or difficult they thought it was to address the criteria, how prepared they felt to address the criteria, how they approached addressing the criteria, whether they thought the criteria were sufficient, and what, if any, additional criteria they would recommend. Participants also discussed the extent to which they thought their repositories were more secure as a result of adhering to these criteria. Appendix A includes the

interview protocol. The interviews lasted approximately 30 minutes and took place on Skype.

All interviews were audio-recorded and transcribed. Afterwards, transcripts were coded using NVivo – a qualitative data analysis software tool. Prior to analyzing the transcripts, we developed a codebook based primarily on the three main dimensions of security established in the computer science literature: confidentiality, integrity, and availability. Specifically, two members of the research team coded the transcripts looking for any statements participants made that corresponded to the concepts of confidentiality, integrity, and availability. We then calculated a table enumerating the frequencies with which participants mentioned each concept in relation to their perceptions. Finally, we calculated inter-rater reliability using Cohen’s kappa, achieving a score of 0.79.

### 3.2 Surveys

In developing our survey, we examined the literature on security in digital preservation and computer science, including research on security metrics. We did not find an existing instrument to measure the security perceptions of computing infrastructures by those who are responsible for managing and securing said infrastructure. Consequently, we derived items for our survey from definitions and explanations of confidentiality, integrity, and availability in Bishop [1], a foundational text on computer security.

We asked the same individuals that we interviewed to take our survey. The survey consisted of 19 items: 4 pertaining to confidentiality, 11 pertaining to integrity, and 4 pertaining to availability. The survey included a 5-point, likert-type scale ranging from “Strongly disagree” to “Strongly agree” with one additional option: “Not applicable.” Appendix B includes the survey instrument. The items were randomized to mitigate order effects.

To analyze the survey data, we calculated descriptive statistics, including participants’ mean scores on the items that pertained to confidentiality, integrity, and availability. We also performed the Kruskal-Wallis H test to identify whether there were any statistically significant differences in participants’ attitudes toward the confidentiality, integrity, and availability principles.

## 4. FINDINGS

The findings are organized based on the methods we used to collect the data. After discussing participant characteristics, we discuss findings from the interviews. Next, we discuss findings from the surveys.

### 4.1 Participant Characteristics

Two people participated in this study, one from each organization that successfully acquired the nestor seal of approval. Both participants held senior positions in the organizations where they worked. Their responsibilities included overseeing teams involved in national and international digital preservation projects and initiatives as well as policy and services development within their organizations. Participants reported working approximately five to nine years on digital repositories at their current organizations. Both

participants reported having involvement in the development of standards for digital repositories.

### 4.2 Interview Findings

Participants shared their views on the concept of security for digital repositories. Specifically, they viewed security as a prerequisite for trustworthiness. They saw security as making sure that repositories act as they are supposed to with no intended or unintended interruptions.

Participants also shared their views on criterion C34 and its explanatory notes. They thought that criterion C34 itself was a bit general, but the explanatory notes for C34 were a helpful complement, providing guidance on how to successfully address the criterion within their repositories. Despite the fact that participants found it difficult to address criterion C34, they felt prepared to address it based on the security measures they had in place prior to audit (e.g., redundant storage, protection against data manipulation, and implementation of national IT standards). While participants did not consider their repositories more or less secure as a result of addressing the explanatory notes for criterion C34, they thought their documentation for what they do to secure systems improved. When asked whether the explanatory notes for criterion C34 set the bar for security too high, too low, or just right, participants stated that addressing the explanatory notes sets the bar just right, suggesting that they considered the security requirements for nestor certification as reasonable, appropriate, and sufficient for securing their repositories.

Analysis of interview data against the confidentiality, integrity, and availability security principles established in computer science revealed that participants provided statements pertaining to the concept of integrity most frequently, followed by availability and confidentiality. Table 1 lists the frequency with which participants provided statements pertaining to each concept. When participants mentioned integrity, they referred to protecting their data from any threats, including manipulation. Participants mentioned the importance of confidentiality and availability because both are included in the nestor definition of security—a definition which they reported as being important to their work. They did not, however, elaborate on what either of the concepts meant to them in their practice.

**Table 1. Frequency Participants Mentioned Security Concepts**

Security Concepts	Frequency
Confidentiality	2
Integrity	10
Availability	2

### 4.3 Survey Findings

To complement the interview data and get a better sense of the relevance of security principles to the participants, we administered surveys to them. The surveys asked questions about participants’ views on aspects of confidentiality, integrity, and availability.

Table 2 lists the mean scores of participants’ responses for the questions pertaining to each security principle. Comparing the mean scores of participants’ responses to the survey questions

reveals that participants are most concerned with integrity, followed by availability and confidentiality.

**Table 2. Mean Scores for Security Concepts**

Security Concepts	Mean Scores
Confidentiality	3.38
Integrity	4.55
Availability	3.75

A Kruskal-Wallis H test showed that there was a statistically significant difference in participants’ ratings of security survey items based on the different principles the items referred to,  $\chi^2(2) = 7.82, p = .02$ , with a mean rank security score of 13.75 for confidentiality, 23.50 for integrity, and 14.25 for availability. These results suggest that participants had stronger attitudes about integrity relative to their attitudes about availability and confidentiality.

## 5. DISCUSSION

Results underscore the importance of security to the digital repository staff members who participated in this study. Participants mentioned the three security principles of confidentiality, integrity, and availability during the interviews. Participants also rated survey items pertaining to those three principles highly, suggesting that they are relevant to their views on securing digital repositories.

Although participants mentioned the three security principles of confidentiality, integrity, and availability during the interviews, and rated survey items pertaining to them highly, results of this study provide more empirical support for some principles of security than others. For example, participants provided more statements related to integrity than availability and confidentiality. As another example, participants rated survey items pertaining to integrity higher than survey items pertaining to availability and confidentiality. The fact that the interview data and survey data triangulate with respect to more emphasis on integrity relative to availability and confidentiality is interesting and needs to be looked at more in depth in future research. The main questions that we need to understand going forward are: Why is integrity more salient to digital repository staff members? And what might this mean for research and practice? First, we need to understand whether having more questions pertaining to the concept of integrity has an effect on the results. Second, we need to understand whether we would still receive more empirical support for integrity than availability or confidentiality if a similar study was conducted with a larger sample of participants. This would enable us to know if the study participants’ views on security generalize to other digital repository staff members. Third, we need to understand what impact digital repository staff members’ views on security actually have on the security of digital repositories. For example, if the principle of integrity is more salient in digital repository staff members’ minds, does this mean that digital repositories are less secure when it comes to availability and confidentiality? In other words, are digital repository staff members focusing on integrity at the expense of availability or confidentiality? This may not be the case. It could simply be that integrity is more important than availability or confidentiality. Or it could be that performing actions related to integrity indirectly address issues relating to availability and confidentiality. Or it could be that digital repository managers

find it easier to address availability and confidentiality relative to integrity, and so they focus on integrity. At any rate, future research should seek to address these issues so that we can have a better understanding of how what digital repository staff members think about security affects the security of digital repositories.

This study makes two primary contributions to the digital preservation literature. First, it complements the development of standards for TDRs by focusing on the security criteria within one of those standards – DIN 31644. This study examines these security criteria from digital repository staff members’ points of view. Prior to this study, we only had the security criteria without insight into the perspectives of those who are responsible for actually addressing those criteria. Second, this study also contributes to the digital preservation literature by providing both qualitative and quantitative data collection instruments which can be used to understand digital repository staff members’ perceptions on security. Since efforts to certify trustworthy digital repositories are well underway, and security is a critical element of becoming certified, we anticipate that better understanding digital repository staff members’ perspectives on security will only increase in importance going forward.

This study also makes one main contribution to the computer science literature pertaining to security. It takes a classic definition of security, one underpinned by the principles of confidentiality, integrity, and availability, and moves that definition forward by operationalizing the concept with measurement items in a survey instrument. This instrument, what we call the Security Perception Survey (SPS), represents a security metric focused on the perceptions of those responsible for managing and securing computing infrastructures. While SPS was developed using the responses of people who manage and secure TDRs, one specific type of computing infrastructure, subsequent studies could assess the generalizability of SPS to provide insights into the perceptions of people who are responsible for managing and securing other types of computing infrastructures.

The primary limitation of this study is its sample size. Only two digital repository staff members participated in this study. Thus, we cannot generalize the results of this study beyond our sample. However, we felt that who participated in this study was more important than how many. We needed individuals who were at organizations where third parties had verified the success of their security efforts. We felt these individuals would provide the most insightful information about their views on security. We also thought that staff at organizations that successfully passed repository certification by the DIN 31644 standard would be in the best position to evaluate the security criteria within the standard. These issues guided our choices regarding who was eligible to participate in our study, which in turn, led to a small sample size. Despite our small sample size, we reached 100% of our sampling frame; representatives from all of the organizations that have acquired nestor seals of approval participated in this study. It is also important to note the challenges to employing traditional research methods, such as interviews and surveys, to study security. For example, people are reluctant to participate in security studies because: 1) they have concerns about whether the information they provide could somehow be used by others to compromise their systems, or 2) they fear their own shortcomings with respect to their expertise might become exposed as a result of participation [18]. Although we faced a number of these well-documented

challenges to recruiting participants for our study, we were yet able to successfully recruit individuals from both organizations that recently acquired nestor seals of approval.

## 6. CONCLUSION

Security is a major issue for digital repositories. Digital repository staff members are responsible for managing and securing digital repositories, thus their perspectives on security are critically important to understand. This study provided a preliminary investigation into digital repository staff members' views on security and security criteria in standards for TDRs, in particular DIN 31644 and the nestor explanatory notes for Trustworthy Digital Archives. Participants articulated their views on security in terms of integrity and to a lesser extent availability and confidentiality. Results of this study warrant a closer correspondence between research on security in digital preservation and computer science, because of the overlap that results of this study have demonstrated. Participants in this study found the security criteria in the standard that they chose sufficient. Going forward, researchers should continue analyzing digital repository staff members' views on security and security criteria, so that the digital preservation community can validate the relevance and importance of the security criteria by those who are responsible for making digital repositories secure.

## 7. ACKNOWLEDGMENTS

We thank Michael Frisby and his colleagues at the Indiana Statistical Consulting Center for providing assistance with data analysis and reading prior drafts of this work.

## 8. APPENDICES

### 8.1 Appendix A – Interview Protocol

1. How do you define repository trustworthiness? In other words, what does it mean to you for a repository to be trustworthy?
2. How do you define security as it relates to digital repositories? In other words, what does security mean to you?
3. How would you describe the relationship between the trustworthiness of a digital repository and the security of that digital repository? In other words, how would you describe the relationship between security and trustworthiness?
4. Take a minute to read over C34, the nestor criterion on security. Now think back to when you were preparing for audit. How easy or difficult was it to address criterion C34 for your digital repository?
5. How much time do you think it took you and your colleagues to address criterion C34?
6. How prepared were you and your colleagues to address criterion C34?
7. Do you think your repository is more secure as a result of addressing criterion C34? Why or why not?
8. Do you think criterion C34 sets the bar too high for addressing security issues? Or do you think criterion C34 sets the bar too low for addressing security issues? Or do you think criterion C34 sets the bar "just right" for addressing security issues? Why or why not?

9. Do you think any additional criteria should be added to criterion C34 to make digital repositories more secure and therefore more trustworthy? If so, how would you describe what criteria should be added?
10. Did you use DRAMBORA to help you address the security criteria in DIN 31644? If so, which parts of DRAMBORA were most helpful and why?
11. Is there anything else you'd like to add, given our topic of security of Trustworthy Digital Repositories?

### 8.2 Appendix B – Security Perceptions Survey

Questions pertaining to confidentiality (Questions were answered on a 5-point, likert-type scale ranging from "Strongly disagree" to "Strongly agree" with one additional option: "Not applicable.")

1. Access control mechanisms should be used to support confidentiality (e.g., cryptography).
2. Mechanisms should be used to prevent illicit access to information.
3. The existence of data should be denied to protect it.
4. Resources should be hidden to protect them.

Questions pertaining to integrity (Questions were answered on a 5-point, likert-type scale ranging from "Strongly disagree" to "Strongly agree" with one additional option: "Not applicable.")

1. Improper changes to data should be prevented.
2. Unauthorized changes to data should be prevented.
3. Information about the source of data should be protected.
4. Unauthorized changes to information about the source of data should be prevented.
5. Prevention mechanisms should be used to maintain the integrity of data by blocking any unauthorized attempts to change the data.
6. Prevention mechanisms should be used to maintain the integrity of data by blocking any attempts to change the data in unauthorized ways.
7. Detection mechanisms should be used to report when the data's integrity is no longer trustworthy.
8. System events (e.g., user or system actions) should be analyzed to detect problems.
9. The data itself should be analyzed to see if it has been changed.
10. A system should report what causes integrity violations.
11. A system should report when a file is corrupt.

Questions pertaining to availability (Questions were answered on a 5-point, likert-type scale ranging from "Strongly disagree" to "Strongly agree" with one additional option: "Not applicable.")

1. A system should guard against denial of data attacks.
2. A system should guard against denial of service attacks.
3. An unavailable system is at least as bad as no system at all.

4. A system administrator should be able to tell the difference between when data is not available due to circumstances in the environment versus a security attack.

## 9. REFERENCES

- [1] Bishop, M., 2003. *Computer security : art and science*. Addison-Wesley, Boston.
- [2] Butler, S.A. and Fischbeck, P., 2002. Multi-attribute risk assessment. In *Symposium on Requirements Engineering for Information Security*. <http://openstorage.gunadarma.ac.id/research/files/Forensic/OpenSource-Forensic/MultiAttributeRiskAssesment.pdf>
- [3] Cherdantseva, Y. and Hilton, J., 2013. A reference model of information assurance & security. In *Availability, reliability and security (ares), 2013 eighth international conference on IEEE*, 546-555. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6657288&isnumber=6657192>
- [4] Deutsches Institut Für Normung, 2012. *Information and documentation—criteria for trustworthy digital archives*. Deutsches Institut für Normung. <http://www.din.de/en/getting-involved/standards-committees/nid/standards/wdc-beuth:din21:147058907>
- [5] Digital Curation Centre and Digital Preservation Europe, 2007. *DCC and DPE Digital Repository Audit Method Based on Risk Assessment, v1.0*. <http://www.repositoryaudit.eu>
- [6] Dillo, I. and De Leeuw, L., 2014. *Data Seal of Approval: Certification for sustainable and trusted data repositories*. Data Archiving and Networked Services. [http://www.datasealofapproval.org/media/filer\\_public/2014/10/03/20141003\\_dsa\\_overview\\_defweb.pdf](http://www.datasealofapproval.org/media/filer_public/2014/10/03/20141003_dsa_overview_defweb.pdf)
- [7] Herrmann, D.S., 2007. *Complete guide to security and privacy metrics : measuring regulatory compliance, operational resilience, and ROI*. Auerbach Publications, Boca Raton.
- [8] Howard, M. and Leblanc, D.E., 2002. *Writing Secure Code*. Microsoft Press.
- [9] International Organization for Standardization, 2012. Space data and information transfer systems: audit and certification of trustworthy digital repositories International Organization for Standardization. [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=56510](http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510)
- [10] Jansen, W., 2009. *Directions in Security Metrics Research*. National Institute of Standards and Technology. [http://csrc.nist.gov/publications/nistir/ir7564/nistir-7564\\_metrics-research.pdf](http://csrc.nist.gov/publications/nistir/ir7564/nistir-7564_metrics-research.pdf)
- [11] Nestor Certification Working Group, 2013. *Explanatory notes on the nestor Seal for Trustworthy Digital Archives*. [http://files.dnb.de/nestor/materialien/nestor\\_mat\\_17\\_eng.pdf](http://files.dnb.de/nestor/materialien/nestor_mat_17_eng.pdf)
- [12] Organization for Economic Co-Operation and Development, 2002. *OECD Guidelines for the Security of Information Systems and Networks: Towards a Culture of Security*. Organization for Economic Co-operation and Development. <http://www.oecd.org/sti/ieconomy/15582260.pdf>
- [13] Parker, D.B., 1998. *Fighting computer crime : a new framework for protecting information*. J. Wiley, New York :.
- [14] Pfleeger, S. and Cunningham, R., 2010. Why measuring security is hard. *IEEE Security & Privacy*, 4, 46-54. <http://doi.ieeecomputersociety.org/10.1109/MSP.2010.60>
- [15] Pfleeger, S.L., 2000. Risky business: what we have yet to learn about risk management. *Journal of Systems and Software* 53, 3, 265-273. doi:10.1016/S0164-1212(00)00017-0
- [16] Sen, S. and Samanta, S., 2014. Information security. *International Journal of Innovative Research in Technology*, 1(11), 224-231.
- [17] Stoneburner, G., Hayden, C., and Feringa, A., 2004. *Engineering Principles for Information Technology Security (A Baseline for Achieving Security), Revision A*. National Institute of Standards and Technology. <http://csrc.nist.gov/publications/nistpubs/800-27A/SP800-27-RevA.pdf>
- [18] Sundaramurthy, S.C., Mchugh, J., Ou, X.S., Rajagopalan, S.R., and Wesch, M., 2014. An anthropological approach to studying CSIRTs. *IEEE Security & Privacy*, 5, 52-60. doi:10.1109/MSP.2014.84