The FAIR Accessor and the authenticity of digital archival information

O "FAIR Accessor" e a autenticidade da informação arquivística digital

André Pereira PACHECO[1] 0000-0002-1810-4866

Maria Cristina Vieira de FREITAS[2] 0000-0002-8849-8792

Carlos Guardado da SILVA[3] 0000-0003-1490-8709

Abstract

The constant increase in digital information's volume, variety, and complexity poses many problems that make it difficult to preserve archival information while ensuring that it remains authentic, reliable, accessible, trustworthy, intelligible, and reusable for as long as possible. This study explores the concepts of a possible implementation of a FAIR Accessor, a technology developed to provide Findable, Accessible, Interoperable, and Reusable research data, as an infrastructure that can support and aid archival information description and ensure its authenticity. A qualitative literature review on a selection of representative works in the fields of Information Science, Diplomatics, and the FAIR principles is followed by a discussion on how the key concepts of each field overlay and thus may complement each other mutually. It is concluded that the infrastructure of the FAIR Accessor can prove useful in enriching archival description and, ultimately, in assisting to ascertaining the authenticity of records.

Keywords: Archival description. Digital information. Digital preservation. Diplomatics. FAIR data.

Resumo

O crescimento constante no volume, variedade e complexidade da informação digital coloca uma série de problemas que dificultam a preservação da informação digital ao mesmo tempo que se procura garantir que essa permaneça autêntica, fidedigna, acessível, confiável, inteligível e reutilizável pelo tempo necessário. Este estudo explora os conceitos de uma possível implementação de um "FAIR Accessor", uma tecnologia desenvolvida com o objetivo de criar dados de investigação pesquisáveis, acessíveis, interoperáveis e reutilizáveisAIR). Tratase de uma infraestrutura capaz de apoiar a descrição arquivística e de garantir a autenticidade da informação. A investigação consiste numa revisão de literatura qualitativa de obras selecionadas nas áreas da Ciência da Informação e Diplomática e da gestão de dados científicos, em particular nos princípios FAIR, seguida de uma discussão de como certos conceitoschave de cada domínio se sobrepõem e, potencialmente, se complementam mutuamente. Concluise que a infraestrutura de um "FAIR Accessor" pode revelarse útil em enriquecer a descrição arquivística além de poder contribuir para a presunção da autenticidade dos documentos de arquivo.

Palavras-chave: *Descrição arquivística. Dados FAIR. Diplomática. Preservação digital. Informação digital.*

[1] Universidade de Coimbra, Centro de Estudos Interdisciplinares do Século XX. Coimbra, Portugal. *Corresdpondência para/*Coreresponde nce to: A. P. PACHECO. E-mail: <andrez.pacheco@gmail.com>.

[2] Universidade de Coimbra, Centro de Estudos Interdisciplinares do Século XX. Coimbra, Portugal.

[3] Universidade de Lisboa, Faculdade de Letras, Centro de Estudos Clássicos. Lisboa, Portugal.

Received on March 3, 2021 and approved on November 17, 2021.

Como citar este artigo/How to cite this article

Pacheco, A. P.; Freitas, M. C. V.; Silva, C. G. The FAIR Accessor and the authenticity of digital archival information. *Transinformação*, v.34, e210006, 2022. https://doi.org/10.1590/2318-0889202234e210006

Introduction

There is no denying that the world is more digital by the day. As we further ground our activity in the digital realm, digital records and the procedures that govern their creation emerge as an essential part of commerce, government, research, and citizenship. Emails, online applications, research data, databases, and digital art are but some of the plurality of resources used in the normal operation of modern societies.

The new record formats that appear tend to be more complex, allowing for the creation and manipulation of a wider variety of information representations, such as images, sound, text, video, and the ability to distribute them in unprecedented ways across time and space. However, while technology makes the production, transmission, manipulation, organization, maintenance, and consultation of records easier, faster and cheaper, it also represents a threat to their integrity, accessibility, and preservation. There is a tradeoff between integrity and durability for speed, commodity, and reduced costs.

Thus, it becomes indispensable to preserve the reliability and authenticity of these records to ensure that they constitute trustworthy evidence of the activities they support, whether it is research, legal disputes, copyright, accountability, or individual rights. Proper records management is not a goal but rather a fundamental means of ensuring the protection of rights, bolstering of knowledge, cultural enrichment, and the continuation of memory. This task is critical since, in contrast with traditional records, "[…] digital information is not subject to gradual decay: it either exists, or it does not" (The United Nations Educational, Scientific and Cultural Organization, 2007, p. 2). This means there might not be second opportunities. Time is at the essence.

As a result, we have witnessed considerable losses of digital materials throughout the last few decades due to changing technology and inadequate preservation practices. For those that survived, it is currently difficult to ascertain their authenticity (Duranti, 2007). Even not so long ago, Ferreira, Saraiva e Rodriguez (2012) noted that many repositories do not possess formal preservation policies. Additionally, all the 35 Portuguese repositories studied by the authors failed to showcase a formal preservation policy. In addition, despite the number of existing standards, these are often considered difficult implementation (Duranti; Blanchette, 2004), which often translates into puzzled information professionals and insufficient data preservation actions.

The still existing lack of inadequate practices in addressing the issues of preserving the authenticity of records in the long term, which frequently leads to the inability to constitute evidence, demands further research. Therefore, this study attempts to contribute with a theoretical approach that analyzes the concept and structure of a FAIR Accessor, a resource architecture based on the core principles that research data needs to be Findable, Accessible, Interoperable, and Reusable (FAIR), and discusses its reproducibility within the framework of Information Science as a means of bolstering archival description, specifically for ascertaining a digital record's authenticity. Since the digital preservation of authentic information is a vast and complex process involving access, costs, storage, technological decay, representation, and data transformation, this study focuses merely on access and availability of a record's archival description to determine their authenticity and thus provide one additional possibility of addressing this chimeric challenge.

Methodological Procedures

The capability of making information FAIR is a key competence in information management. However, this crucial capability is still not observed in many institutions due to the lack of understanding of the informational phenomenon and how to address it. In research data management, the focus has been on making research findings

openly available, whereas in archival science, it is necessary to add an authenticity layer. Despite its similar goals in obtaining FAIRness, the approaches of these two fields have remained fairly independent in current research.

Therefore, this paper provides a literature review that unravels contact points between these previously and seemingly distant approaches. Diplomatics' principles regarding the preservation of authentic records are also considered in the literature review to strengthen their connection. The findings from the literature review are summarized into novel exploratory research that bridges the methodologies of these fields by elucidating how the structure of a FAIR Accessor can be helpful for the description of authentic archival information within the framework of archival description.

This qualitative, exploratory, and bibliographic study aligned with these research lines. The qualitative approach shapes the study as reflexive, synthetic, and analytic, according to which research emerges and is constructed during the exploratory phase, without conforming to a rigid research protocol. As a result, rather than describing the object, it seeks to explore and understand it. Consequently, the results boost the understanding of the identified issues in the research and generate new research hypotheses or recommendations for future research.

The bibliography on which this research is based has been gathered via selective and nonexhaustive criteria. In harmony with the qualitative approach, texts are selected based on intentional and nonprobabilistic sampling. Therefore, the collected data exhibits particularities and differences instead of patterns or repetitions. As a consequence of this approach, 35 texts are selected, which have been written and published in a deliberately broad research interval of 21 years (from 1994 to 2020). Regarding frequency, more emphasis is given to works produced in the last decade (19 documents, 54%) concerning works created in earlier decades of the research interval (16 papers, 46%). Additionally, it is worth mentioning that 35% of the sample (13 documents) comprises texts published over the last five years, an observation following the relative currentness of the research topic.

Finally, the criteria for the selection process are based on the pertinence and the possibility to support, clarify or expand concepts and ideas. Another essential criterion lies in the perceived ability of the works to boost new interpretations and contributions, not only for the discussion of present problems but also, and mainly, to be able to shed new light on them and to generate future research.

Results

The object of information science

The advent of the digital environment has brought a significant change in understanding and managing records. In this section, only a few major aspects will be mentioned, without diving into detail, as it would surpass the scope of this study.

At the core of every new challenge lies the notion that a document is no longer a binary reality. The informationmediumtechnology trinity replaces the traditional bidimensional information medium. The practical implication is that to preserve an analog record's information, proper conservation measures on its medium (paper, parchment) would suffice; in the case of digital records, it is also necessary to consider the technological infrastructure that supports them (FringsHessami, 2020).

The fact that information can easily be dissociated from its original support and freely transformed, reproduced, and disseminated also reveals a crucial idea regarding the object of archival science. If, for decades, its object of study was the record, since document and information embodied a single entity so that handling a document meant obtaining all the available information, for that document, today it is increasingly common to

acknowledge that the object is information (Cook, 2007) which gains form and existence through a document, its carrier. Not generic information, but the trustworthy organizational information, created by a specific body as the byproduct of its activity and preserved for as long as necessary, in permanent access. In this sense, archival science is posited as one of the many disciplines that study social information as part of an overarching Information Science (Silva; Ribeiro, 2008).

As a result, the management of digital archival information has to look beyond the individual record scope and consider the information technology system[4] that provides the connective tissue for its creation, handling, and use. Information flows dynamically and interactively, promoting a shift to more systemic management that considers the recordkeeping system. Consequently, it can be found in literature a growing urge to consider the creation context. Terry Cook brilliantly summarizes this change:

> At the heart of the new paradigm is a shift away from viewing records as static physical objects, and towards understanding them as dynamic virtual concepts; a shift away from looking at records as the passive products of human or administrative activity and towards considering records as active agents themselves in the formation of human and organization memory; a shift equally away from seeing the context of records creation resting within stable hierarchical organizations to situating records within fluid horizontal networks of workflow functionality […] from product to process, from structure to function, from archives to archiving, from the record to the record context, from the "natural" residue or passive byproduct of administrative activity to the consciously constructed and actively mediated ‹‹archivalisation›› of social memory (Cook, 2001, p. 4).

Following Cook's view, archival science should shift from analysing individual records to understanding the functions, processes, and transactions that underlie records' creation. A similar approach has been suggested by Frank Upward, who argues that information should be managed in a systemic and integrated manner as part of a continuum that focuses on accessibility and fluidity (Upward, 2000).

The role of archivists is also affected by this conceptual shift. The classic passive records keeper steps aside, allowing the emergence of information managed with active and constant participation in defining the information flow (Ribeiro, 2005; Arnold; van der Walt, 2019). Cook (2001) and MacNeil (2007) argue that this participation should begin before records' creation (*i.e.*, at the very conception of the system) under the risk that, if later, due to lack of proper preservation policies, there will no longer be any records left to be preserved.

Diplomatic authenticity

The body of concepts and principles of Diplomatics can greatly aid in understanding a record's authenticity. In the 17th and 18th centuries, it originally emerged to prove the reliability and authenticity of records. Over the centuries, it has developed into a "[…] very sophisticated system of ideas about the nature of records, their genesis, and composition, their relationships with the actions and persons connected to them, and with their organizational, social, and legal context" (Duranti; Eastwood, 1995, p. 215).

A record is seen as "[…] any document created by a physical or juridical person in the course of practical activity as an instrument and a byproduct of it, where 'created' means made or received and retained ('set aside') to act or for reference" (Duranti, 1997, p. 216). Its characteristics are having a "[…] fixed documentary form, a stable

[4] The information technological system is understood as the "[…] technological platform – environment or physical and logical medium – that supports the creation, processing, circulation, storage, transmission and access to the information that represents the information system" (Pinto, 2009, p. 25). In turn, the information system "[…] is formed by the different types of recorded information […] regardless of the medium (material and technological), according to a structure (creating/receiving body) carried out over time" (Silva, 2006, p. 162).

content, an archival bond with other records either inside or outside the system, and an identifiable context" (InterPARES 3, 2012). At the heart of diplomatics lies the notion that:

> […] all records can be analyzed, understood, and evaluated in terms of a system of formal elements that are universal in their application and decontextualized in nature. The essential assumption of diplomatics is that the context of a document's creation is made manifest in its form can be separated from, and examined independently of, its content (Duranti, 1997, p. 215).

Therefore, a record is a conceptual embodiment of internal and external elements that must be analyzed to assess its authenticity from a diplomatic perspective.

This idea that all records resulting from the same function exhibit the same documentary form is linked to the notion of archival bond, *i.e.*:

> […] the network of relationships that each record has with the records belonging in the same aggregation. The archival bond is originary, because it comes into existence when a record is created […] necessary because it exists for every record […] and determined because it is qualified by the function of the record in the documentary aggregation in which it belongs (Duranti 1997, p. 215).

In other words, it is the defining characteristic of a record that transforms a document into a record, as it represents the organic relationship that a record shares with others of the same system created in the course of the same activity.

The presumption of the record authenticity is rooted in the belief that the rules that govern records' creation will be manifested in its form. Authenticity refers to the "[…] trustworthiness of a record as a record, i.e., the quality of a record that is what it purports to be and that is free from tampering or corruption" (InterPARES3, 2012). To assess it, one must establish its identity and demonstrate its integrity (Duranti; Blanchette, 2004). Identity is related to the "distinguishing character of a record, that is, the attributes of a record that uniquely characterize it and distinguish from other records", while integrity refers to "[…] its wholeness and soundness: a record has integrity when it is complete and uncorrupted in all its essential aspects" (InterPARES 1, 2002, p. 20). Therefore, a record is considered complete if the message it is meant to convey to fulfill its purpose is unaltered.

In summary, authenticity implies that it is possible to demonstrate that a record is what it is meant to be and that its transmission over space and time has not changed it significantly. The term 'significantly' is used since no medium is immune to change, paper decay, or technological obsolescence. However, in electronic media, this process is greatly exacerbated. Since the informational value far outlasts the life expectancy of media, which is five years on average, some form of change is necessary. For this reason, "[…] it is virtually impossible to deliver any preserved electronic record in such a way that none of its elements have changed. To attest the authenticity of a preserved electronic record, then, is to demonstrate that no essential element of the record has changed" (Duranti; Blanchette, 2004, p. 2). In the digital environment, the closest we can get to integrity, by definition a constant state, is a controlled changeability documented by metadata.

In this regard, there is a significant difference between an analog and digital record. While, for the former, to ensure the conservation of its medium is equivalent to safeguarding its integrity and, consequently, its authenticity, for the latter, this no longer holds, but not because the nature of records has changed. Essentially, they remain a complex of elements and their relationships, with many identifiable characteristics, such as a fixed documentary form, stable content, an archival bond with other system records, and are created in the course of action (InterPARES 1, 2002). However, their fundamental difference is that "[…] these components are not inextricably joined the one to the other, as in traditional records: they, and their parts, exist separately, and can be managed separately, unless

they are consciously tied together to ensure the creation of reliable records and the preservation of authentic records" (Duranti; Macneil, 1996, p. 49).

Therefore, the unity of a traditional record in single physical support is now scattered, for digital records, across the recordkeeping system. Likewise, "[…] the contingencies that endow authenticity are observable in the document itself but the procedures of creation, maintenance, and preservation" (Rogers, 2015, p. 20). Classic authentication mechanisms, such as seals, stamps, or signatures, have been shown to actually hamper longterm preservation of digital records, especially digital signatures (InterPARES 1, 2002). Instead, the "[…] best method of ensuring ongoing authenticity of electronic records is external to the records themselves and involves a tight control on recordmaking and recordkeeping procedures" (Duranti; Blanchette, 2004, p. 4). Subsequently, it becomes clear the crucial role of metadata for determining authenticity, understanding the provenance of sources and placing them in a larger context (Graham, 1995; Gilliland, 2016; Zeng; Qin, 2016). This is particularly crucial for digital materials due to the ease in altering, copying and removing them from the original context. This vulnerability calls for an inclusion of metadata for authenticity early on the preservation plan, preferably during the phases of planning and conception of a recordkeeping system (Bountouri; Gratz; Sanmartin, 2017; Rolan, 2017).

Structure of a FAIR Accessor

In January 2014, a range of stakeholders came together in a workshop organized at the Lorentz Centre in Leiden, the Netherlands, to debate the infrastructures that support data science. These discussions led to the definition of a set of guiding principles and practices that could enable data providers and consumers – machines and humans alike – to more easily find, access, interoperate, and reuse the vast quantities of information generated by contemporary dataintensive science. These would be formalized as the FAIR principles, later presented to the academic community by Wilkinson (2016, 2017) and his colleagues, receiving significant attention and importance. For instance, open access has become an obligation in Horizon 2020 for scientific peerreviewed publications (European Commission, 2016).

It is possible to identify many reasons that triggered such debates on existing infrastructures. One factor lies in the vast amounts of data and metadata produced in the course of today's dataintensive science (Corujo; Silva; Revez, 2016; Hemphill; Leonard; Hedstrom, 2020). This boom in value is paired with an increase in the variety of formats and representations, which, by itself, complicates the mission of interoperability, understood as "[…] the ability of data or tools from noncooperating resources to integrate or work together with minimal effort" (Wilkinson *et al.*, 2016, p. 2). To successfully preserve these increasing amounts and variety of data, researchers turn to repositories specific to a research domain (*e.g.*, GenBank, UniProt, Space Physics Data Facility) or generalpurpose (*e.g.*, Figshare, Zenodo, Dataverse, or institutional ones). While the specialpurpose repositories usually possess funds to curate their data and are endorsed with specific query interfaces and rich API to explore their content, generalpurpose ones often allow publication in arbitrary formats, with limited curation and often little structured metadata (Wilkinson *et al.*, 2017). In its way, each poses threats to data discoverability and reuse. The existence of own APIs makes difficult crossrepository queries, while their absence results in the need to manually discover, download and reuse data [30]. As a result, researchers usually need several weeks or even months of technical endeavor to simply collect data for their research, assuming it is successfully obtained. Therefore, the goal of making data FAIR is the creation of "[…] highquality publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies" (Wilkinson *et al.*, 2016, p. 1).

The FAIR principles are theoretical, a guideline that should be followed, but can be materialized in several ways. They define characteristics that contemporary data resources, tools, vocabularies, and infrastructures should exhibit to assist discovery and reuse by third parties. Wilkinson and his colleagues propose an implementation structure based on the combination of three existing technologies: the World Wide Web's Consortium's (W3C)

Linked Data Platform (LDP) (W3C, 2015) the Resource Description Framework (RDF) Mapping Language (Dimou *et al.*, 2014), and the Triple Pattern Fragments (TPF) (Verborgh *et al.*, 2016). Their combined use provides the technological infrastructure necessary for developing three tools, the FAIR Accessor, the FAIR Projector, and the FAIR Profile. Only the FAIR Accessor will be analyzed and discussed in the following lines. It is the basic layer of a resource's representation, therefore the most conceptually similar to an archival description. The other two are more prone to interoperability, which falls outside the scope of this research.

So, what does it exactly mean to be FAIR? In short, the principles state:

• Findable: data should be identified using globally unique, resolvable, and persistent identifiers and should include machineactionable contextual information that can be indexed to support human and machine discovery of that data.

• Accessible: identified data should be accessible, optimally by both humans and machines, using a clearlydefined protocol and, if necessary, with clearlydefined rules for authorization/authentication.

• Interoperable: data becomes interoperable when it is machineactionable, using shared vocabularies and/or ontologies, inside of a syntactically and semantically machineaccessible format.

• Reusable: Reusable data will first be compliant with the F, A, and I principles, but further, will be sufficiently welldescribed with, for example, contextual information so that it can be accurately linked or integrated, with other data sources. Moreover, there should be sufficiently rich provenance information to properly cite reused data (Wilkinson *et al.*, 2017).

To comply with these criteria, Wilkinson and his colleagues started by developing the FAIR Accessor (Figure 1), a lightweight HTTP Interface (Fielding; Taylor, 2002), as it provides "[…] resolvable identifiers for all entities, and a common machineaccessible approach to discovering and retrieving different representations of those entities" (Wilkinson *et al.*, 2017, p. 6). Additionally, the two resource types used in the FAIR Accessor utilize the concept of the Linked Data Platform Container. The LDP "[…] defines a set of rules for HTTP operations on Web resources […] to provide an architecture for readwrite Linked Data on the Web" (W3C, 2015). All entities are identified by a URI, with machinereadable metadata describing the function or purpose of each URI and the nature of the resource that will be returned when that URI is resolved (Wilkinson *et al.*, 2017), and data can be recovered using the HTTP Protocol.

The first type, the "Container Resource", represents a "metadata document that describes the shared features of a collection of resources, and (optionally) the membership of that collection". The second, the "MetaRecord" resource, describes "[…] a member of the contained collection and (optionally) provides ways to access the record itself" (Wilkinson *et al.*, 2017, p. 8).

The Container Resource has the particularities of describing any research object (either a repository, a database, a workflow, a record, and so on). Its representation could include any information, such as authorship/ownership of the object, version number, or preservation policy. The Container URL provides a resolvable identifier (a link) independent from the identifier of the dataset being described, which also means that it may or may not include further URLs representing the MetaRecord Resources. Finally, they can be published freely, *i.e.*, anyone can publish metadata about any research object.

The MetaRecord resource is a specific element within a collection (a record, a study, and so on). Its representation should include rich descriptive and contextual metadata, such as licensing and accessibility and citation information. It may include further URLs that provide direct access to the data itself, with a clear mention of its extension, according to the MIME type (*e.g.*, text/html, image/bmp, audio/wav, amongst many others). Similar to the Container Resources, these may also be published by anyone.

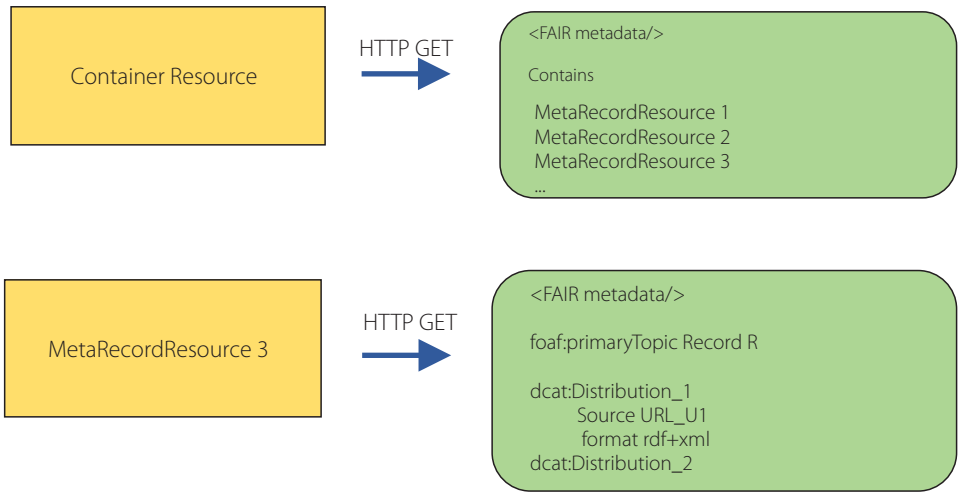

Figure 1. The two layers of the FAIR Accessor.

Note: The authors provide a further explanation in the caption in their paper: "[…] inspired by the LDP Container, there are two resources in the FAIR Accessor. The first resource is a Container, which responses to an HTTP GET request by providing FAIR metadata about a composite research object, and optionally a list of URLs representing MetaRecords that describe individual components within the collection. The MetaRecord resources resolve by HTTP GET to documents containing metadata about an individual data component and, optionally, a set of links structured as DCAT Distributions that lead to various representations of that data" (Wilkinson *et al*., 2017, p. 8).

Source: Adapted with permission from Wilkinson *et al.* (2017, p. 8).

In short, the FAIR Accessor is characterized by the inclusion of rich metadata that facilitates discovery and interoperability of both repository and recordlevel information, described by widely accepted vocabularies. Its two levels are independent, allowing their access and distribution separately. This could be helpful, for instance, for security or access restriction purposes. In addition, each resource type has a unique identifier (URL) with direct reference to higher and lower description levels, enabling access and discoverability from any point of the structure. Another key feature is the absence of an API. It does not require additional technology development; it becomes a lowcost implementation that allows any webcrawler agent to discover the data, acting as a web page. Also, the Accessor is readonly, as it uses the HTTP Get protocol for machines and humans to resolve the unique identifiers and explore its metadata.

In conclusion, it can be envisioned as "[…] a series of Web pages, each containing metadata, and hyperlinks to more detailed metadata and/or data, where the metadata elements and relationships between the pages are explicitly explained to Web crawlers" (Wilkinson *et al*., 2017, p. 9).

Discussion

In the previous sections, we revisited the key tenets of archival information management, of diplomatic principles to ensure a record's authenticity, and finally, of the architecture of a FAIR Accessor for research data management. A discussion will follow of how this illustrative implementation can serve the interests of records management regarding ensuring the authenticity of digital information. Based on the notions revealed by the literature review, it is possible to formulate the following desiderata of the characteristics of digital records regarding their authentic preservation, access and description.

A) Authentic. It includes:

A1) Identity: descriptive metadata for content, structure, and context.

https://doi.org/10.1590/2318-0889202234e210006

A2) Integrity: documentation of changes made to the record.

B) Archival bond: explicit the organic relations that a record has with other records in the system created during the same action or procedure.

C) Permanent access: data needs to be continuously available.

C1) Each record possesses a unique identifier.

C2) The host location is stable and longlasting.

C3) Information is recorded in interoperable formats that allow continuous interpretation.

D) Access granularity: confidential or private information must be protected from unauthorized access.

E) Economic: costs must be kept low to maximize preservation efforts within an almost always limited budget.

The FAIR Accessor can assist in boosting archival description according to an infrastructure that allows fulfilling these objectives. Regarding (A1), either layer of the Accessor provides a proper and predictable place to include every metadata considered relevant, whether at the aggregate level (*e.g.*, information about custody or administrative history) or the itemlevel (*e.g.*, author, title, creator, medium, and so on). The choice of these fields could and should comply with international archival description standards, such as ISAD(G) and metadata standards (*e.g.*, Dublin Core, PREMIS, METS). The integrity of records (A2) can be determined by providing, for example, a .xml (a universal language) distribution containing a registry of all documented alterations, and/or by indicating in the header the number of times it was altered, by who, and what was altered.

The archival bond (B) is guaranteed, on the one hand, by the fact that each resource level possesses an indication of its place in the hierarchy, *i.e.*, each layer of the accessor exhibits information on the resources it contains/is contained in, therefore connecting a record to others that are functionally related to it. On the other hand, the rich descriptive metadata also helps identify the bond (*e.g.*, the creator(s) of the collection).

The structure is webbased regarding permanent access (C) and does not possess any API for data exploration. This means that resources are permanently available and discoverable on the web, both humans and machines. Therefore, (C1) is resolved by defining a unique identifiable URL for each record that can serve as its access point. A permanent link assures the stability of the host location (C2). The interoperability of records' formats (C3) can be achieved using other technologies (TPF and RDF) not explained in this paper. As far as archival description is regarded, interoperability can be ensured by using the .xml format to store any pertinent metadata.

The granularity of access (D) lies in the fact that the two layers of the FAIR Accessor are not interdependent; one can exist without the other. In this sense, it is possible to access a record's metadata without including a distribution link to the data itself. This can be particularly useful to limit access to confidential or sensitive information in archival terms. In addition, the separation of a record and its metadata, in terms of description, allows for interesting preservation scenarios. For instance, when a record is destroyed, the metadata of that record might remain, and it is desired that it does so to enhance accountability. Likewise, if a record, due to technological obsolescence, is no longer available/readable, at least its descriptive metadata is (as it is stored in a universal format, .xml), which enhances preservation mechanisms.

Finally, the lowcost (E) derives from the inexistence of the need to invest in acquiring any new records management software, which is usually costly. Mark Wilkinson and his colleagues suggested that the FAIR implementation infrastructure is APIfree, based solely on available technologies. The single potential cost would be to invest in human resources training.

Table 1 summarizes the major explanations about the potential of integration of FAIR Accessor and archival description model in publishing fair data.

Table 1. Integration of FAIR Accessor and archival description model.

Conceptual dimensions/subdimensions	FAIR Accessor alignments
A. Authenticity	
A1. Identity	Fair Accessor provides a proper and predictable place to include all relevant metadata.
A2. Integrity	Fair Accessor provides a .xml distribution registry of all recorded alterations.
B. Archival bond	Fair Accessor exhibits information on the resources it contains/is contained in.
C. Permanent access	
C1. Unique identifier	Fair Accessor is aligned with defining a unique and stable identifiable URL for each record that can serve as its access point.
C2. Stable host location	Fair Accessor is aligned with assuring permanent links for the resources along time.
C3. Interoperable formats	Fair Accessor is aligned with specific technologies (TPF and RDF) or the use of the .xml format to store selected metadata.
D. Access granularity	FAIR Accessor layers of data and metadata are independent.
E. Economic	FAIR Accessor infrastructure is API-free. Potential costs must include the training of human resources.

Source: original, 2021.

Conclusion

Every passing day and year, the amount of information produced increases radically[5]. This surge in volume is also followed by a greater variety of information representations, mediums, and formats, which translates into an evergrowing complexity of records that emerge as the byproducts of our activities, and that is necessary to preserve for evidence of rights, citizenship, research, and cultural heritage. It is up to archivists and other information professionals to ensure that this sea of information is preserved to remain accessible, authentic, findable, intelligible, trustworthy, and usable for as long as necessary.

However, the mission of trying to preserve in an as unaltered state as possible information which exists in a volatile and changing digital environment is often a truly Sisyphean task. Financial, technological and social problems pose serious obstacles to records' preservation. In addition, there is not a single allpurpose preservation strategy, depending on the choice heavily on contextual needs. However, archival description, as a means of elucidating the nature of the archival bond in its documentary context, is most likely the best method of ensuring records' authenticity in the long term. In this sense, an implementation of the FAIR principles, the FAIR Accessor, was subject to analysis.

The goal was to explore a new method, originated in research data description and management, to describe digital archival information in a way that promotes its authentic preservation, and ensures that it is findable, accessible, and reusable. This study reveals that such a contribution is possible and potentially significant. The FAIR Accessor was shown to provide an infrastructure that includes rich and comprehensive metadata; preserves and makes explicit the archival bond; allows for granularity of access; makes metadata permanently available, even when the record itself is not; allows for both machine and human discoverability, improving the power of queries; has universal access; provides unique identifiers, and has very low implementation requirements.

In conclusion, this technology can potently tackle the highly complex issue of preserving authentic records by complementing existing strategies. Yet to be done and researched are, for instance, an experiment that creates a FAIR Accessor, according to archival criteria. Also, the FAIR Projector and FAIR Profile, which were not covered in the scope of this study, require further research to study their viability from an archival perspective. After all, some developments in digital preservation remain largely experimental, and the history of recent decades does not leave

[5] In 2010, at the Techonomy Conference in Lake Tahoe, Eric Schmidt estimated that every two days we produce as much data as we produced in all history until 2003. This number can only have significantly increased until today, in 2018.

much room for optimism. Nevertheless, we have to keep pushing forward, even if by trial and error so that, in the wise words of Duranti (2007, p. 11): "[…] today's actions, thoughts, achievements and creations will have a future, and the future will have a memory".

Acknowledgments

This study was partially supported by *Fundação para a Ciência e Tecnologia* (FCT), who sponsors André Pacheco under the PhD research grant SFRH/BD/131004/2017.

Contributions

All three authors were involved in the conception of the paper. A. Pacheco was chiefly responsible for writing, whereas C. G. Silva and M. C. Freitas contributed with methodological design and approach, data analysis, scientific revision, and bibliographical references.

References

Arnold, L.; van der Walt, T. Towards a better description of genre archives: a case study of travel archives. *Cataloging and Classification Quarterly*, v. 57, n. 6, p. 369392, 2019. Doi: https://doi.org/10.1080/01639374.2019.1656693.

Bountouri, L.; Gratz, P.; Sanmartin, F. *Digital preservation*: how to be trustworthy. Olimje: Springer International Publishing, 2017. p. 364374. Available from: http://link.springer.com/10.1007/9783319758268. Accessed on: Mar.12, 2021.

Cook, T. Archival science and postmodernism: new formulations for old concepts. *Archival Science*, v. 1, p. 324, 2001. Doi: https://doi.org/10.1007/BF02435636.

Cook, T. Electronic records, paper minds: the revolution in Information management and archives in the postcustodial and postmodernist era. *Archives & Social Studies*: *A Journal of Interdisciplinary Research*, v. 1, n. 0, p. 399443, 2007. Available from: https://archivo.cartagena.es/doc/Archivos_Social_Studies/Vol1_n0/06cook_electronic.pdf. Accessed on: Mar. 12, 2021.

Corujo, L.; Silva, C. G.; Revez, J. Digital curation and costs: approaches and perceptions. *In*: TEEM '16: Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality, 2016, Salamanca. *Proceedings Online* […]. Salamanca: Facultad de Biblioteconomía y Documentación de la Universidad de Salamanca, 2016. p. 277284. Doi: https://doi.org/10.1145/3012430.3012529.

Dimou, A. *et al.* RML: a generic language for integrated RDF mappings of heterogeneous data. *In*: Proceedings of the 7th Workshop on Linked Data on the Web, 2014, Seoul. *Prceedings Online* […]. Seoul: CEUR, 2014. Available from: http://events.linkeddata.org/ldow2014/papers/ldow2014_paper_01.pdf. Accessed on: Mar. 12, 2021.

Duranti, L. The Archival Bond. *Archives and Museum Informatics*, v. 11, n. 3, p. 213218, 1997. Doi: http://doi.org/10.1023/A:1009025127463.

Duranti, L. *Theoretical Elaborations into Archival Management in Canada (TEAM Canada)*: Implementing the theory of preservation of authentic records in digital systems in small and mediumsized archival organizations. [*S.l.*:*s.n.*], 2007. SSHRC CURA InterPARES 3 Project Proposal. Available from: http://www.interpares.org/display_file.cfm?doc=ip3_detailed_project_proposal_v10.pdf. Accessed on: Mar. 12, 2021

Duranti, L.; Blanchette, J. The authenticity of electronic records: the InterPARES approach. *In*: Proceedings of IS&T 2004 Archiving Conference, 2004, San Antonio. *Proceedings Online* […]. San Antonio: IS&T, 2004. p. 215220. Available from: http://www.interpares.org/display_file.cfm?doc=ip12_dissemination_cpr_duranti~blanchette_ist_2004.pdf. Accessed on: Mar. 12, 2021.

Duranti, L.; Eastwood, T. Protecting electronic evidence: a progress report on research study and its methodology. *Archivi & Computer*, n. 3, p. 213250, 1995.

Duranti, L.; MacNeil, H. The protection of the integrity of electronic records: an overview of the UBCMAS Research Project. *Archivaria*, v. 42, p. 4667, 1996. Available from: https://archivaria.ca/index.php/archivaria/article/view/12153/13158. Accessed on: Mar. 12, 2021.

European Commission. DirectorateGeneral for Research & Innovation. *Guidelines on FAIR data management in Horizon 2020*. Brussels: European Commission, 2016. Available from: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020hioadatamgt_en.pdf. Accessed on: Mar. 12, 2021.

Ferreira, M.; Saraiva, R.; Rodrigues, E. *Estado da arte em preservação digital*. [*S.l.*]: RCAAP, 2012. Available from: http://repositorium.sdum.uminho.pt/handle/1822/17049. Accessed on: Mar. 12, 2021.

Fielding, R.; Taylor, R. N. Principled design of the modern Web architecture. *ACM Transactions on Internet Technology*, v. 2,

n. 2, p. 115150, 2002. Available from: http://portal.acm.org/citation.cfm?doid=514183.514185. Accessed on: Mar. 12, 2021.

FringsHessami, V. The societal embeddedness of records: teaching the meaning of the fourth dimension of the Records Continuum Model in different cultural contexts. *Archival Science*, v. 21, p. 139154, 2020. Doi: https://doi.org/10.1007/s10502020093496.

Gilliland, A. Setting the stage. *In*: Baca, M. (org.). *Introduction to Metadata*. 3rd ed. Los Angeles: Getty Research Institute, 2016. p. 120.

Graham, P. S. Requirements for the digital research library. *College & Research Libraries*, v. 56, n. 4, p. 331339, 1995. Available from: http://crl.acrl.org/index.php/crl/article/view/14982. Accessed on: Mar. 12, 2021.

Hemphill, L.; Leonard, S. H.; Hedstrom, M. Saving social media data: understanding data management practices among social media researchers and their implications for archives. *Journal of the Association for Information Science and Technology*, v. 72, n. 1, p. 97109, 2020. Doi: https://doi.org/10.1002/asi.24368.

MacNeil, H. Archival theory and practice: between two paradigms. *Archivaria*, n. 37, p. 620, 2007. Available from: https://archivaria.ca/index.php/archivaria/article/view/11982. Accessed on: Mar. 12, 2021.

Pinto, M. M. *PRESERVMAP*: um roteiro da preservação na era digital. Porto: Edições Afrontamento, 2009.

Ribeiro, F. O perfil profissional do arquivista na sociedade da informação. *Trabalhos de Antropologia e Etnologia*, v. 45, n. 1, p. 4957, 2005. Available from: http://ler.letras.up.pt/uploads/ficheiros/artigo8871.pdf. Accessed on: Mar. 12, 2021.

Rogers, C. *Virtual authenticity*: authenticity of digital records from theory to practice. 2015. 231 f. Thesis (Doctorate in Library, Archival and Information Studies) – University of British Columbia, Vancouver, 2015. Available from: https://open.library.ubc.ca/clRcle/collections/ubctheses/24/items/1.0166169. Accessed on: Mar. 12, 2021.

Rolan, G. Towards interoperable recordkeeping systems: a metamodel for recordkeeping metadata. *Records Management Journal*, v. 27, n. 2, p. 125148, 2017. Doi: https://doi.org/10.1108/RMJ0920160027.

Silva, A. M. *A informação*: da compreensão do fenómeno e construção do objeto científico. Porto: Edições Afrontamento, 2006.

Silva, A. M.; Ribeiro, F. *Das "ciências" documentais à ciência da informação*: ensaio epistemológico para um novo modelo curricular. 2. ed. Porto: Edições Afrontamento, 2008.

InterPARES1. *Authenticity task force report*: the Longterm preservation of authentic electronic records: Findings of the InterPARES Project. San Miniato: Archilab, 2002. Available from: http://www.interpares.org/book/interpares_book_d_part1.pdf. Accessed on: Mar. 12, 2021.

InterPARES3. Record Characteristic. *In*: Interpares. *The InterPARES Project Dictionary*. [*S.l.*]: Interpares, 2012. Available from: http://www.interpares.org/ip3/ip3_terminology_db.cfm?status=dictionary. Accessed on: Mar. 12, 2021.

InterPARES3. Authenticity. *In*: Interpares. *The InterPARES Project Dictionary*. [*S.l.*]: Interpares, 2012. Available from: http://www.interpares.org/ip3/ip3_terminology_db.cfm?status=dictionary. Accessed on: Mar. 12, 2021.

United Nations Educational, Scientific and Cultural Organization. *Preserving the digital heritage*. Amsterdam: Netherlands National Commission for UNESCO and European Commission on Preservation and Access, 2007. v. 31. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1464905508000146. Accessed on: Mar. 12, 2021.

Upward, F. Modelling the continuum as a paradigm shift in recordkeeping and archiving processes, and beyond – a personal reflection. *Records Management Journal*, v. 10, n. 3, p. 115139, 2000. Doi: https://doi.org/10.1108/EUM0000000007259.

Verborgh, R. *et al.* Triple pattern fragments: a lowcost knowledge graph interface for the Web. *Journal of Web Semantics*, v. 3738, p. 184206, 2016. Doi: https://doi.org/10.1016/j.websem.2016.03.003.

W3C. L*inked Data Platform 1.0*. [*S.l.*]: W3, 2015. Available from: https://www.w3.org/TR/ldp/. Accessed on: Mar. 12, 2021.

Wilkinson, M. e*t al.* The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, v. 3, e160018, 2016. Doi: https://doi.org/10.1038%2Fsdata.2016.18.

Wilkinson, M. *et al.* Interoperability and FAIRness through a novel combination of Web technologies. *PeerJ Computer Science*, v. 3, n. 110, e110, 2017. Doi: https://doi.org/10.7717/peerjcs.110.

Zeng, M. L.; Qin, J. *Metadata*. 2nd ed. London: Facet Publishing, 2016.