A Foundation for Trust in Digital Preservation:
The Preservation as a Service for Trust (PaaST)Project Results

Kenneth Thibodeau

Abstract: Over any time frame that spans multiple generations of information technology, there is uncertainty whether digital objects or outputs from digital objects can be verified as the same things they were originally. The Preservation as a Service for Trust (PaaST)Project of the InterPARES Trust collaboration addressed this challenge with the objective of enabling determination of wether preserved digital information objects remain authentic. This paper, first presented at the InterPARES International Symposium in San Jose, Costa Rica on 19 February 2020,[1] sets out the conceptualization of authenticity that guided the articulation of functional and data requirements to support verifiable preservation and, thus provide a foundation of trust. It then summarizes the services that comprise the requirements and discusses the variety of ways the services can be implemented.

Introduction

The preservation of authentic digital records has an endemic problem of trustworthiness. We cannot guarantee, we cannot even estimate the probability of it being successful. All records are liable to neglect, inadequate controls, poor management, rupture in the chain of custody, and deterioration due to aging, environmental contaminants or intentional assault. Such factors can be addressed and, with records in hard copy, addressing them properly can justify a reasonable expectation of success. With digital records, however, even implementing best practices with regard to such risks on a continuing basis cannot eliminate the uncertainty of success or even reduce it to a comfortable level. That is because the uncertainty does not derive from properties of the records themselves or from their curation, but from the fact that digital records intimately depend on technologies whose future is uncertain and unpredictable. The dependency is intimate because all retrieval and any use of digital records, whether for processing within an automated system or presentation to humans, requires that the bits and bytes that constitute stored digital records must be processed by software. This necessity creates a three-edged sword: (1) the software on which digital records depend originally will become obsolete; (2) the software used to create the records typically also enables their alteration or deletion; (3) software used to retrieve, process or present the digital records may produce output that is not authentic.

These insights came to the fore in the course of the Preservation as a Service for Trust (PaaST) project. PaaST was an initiative of the North American Team in the InterPARES Trust collaboration.[2] It was motivated by issues entailed when digital records are preserved in the Cloud. In the Cloud context, transparency of the

[1] Preservación Digital de Documentos Auteńticos: Investigación y Resultados de InterPARES. Universidad de Costa Rica, San José, Costa Rica. 18-19 Febrero do 2020. http://archivo.ucr.ac.cr/simposio-interpares.htm.

[2] https://interparestrust.org/

technology and technological processes used in preservation is an obvious issue because Cloud service providers generally do not inform their customers of what hardware and software they use, how they are configured or when they change. In the course of addressing technical issues related to preserving records in the Cloud, it was recognized that the three-edged sword weighs over digital preservation scenarios in general. Nonetheless, the initial focus on the Cloud led to the formulation of a distinctive approach, one which did not aim at technological solutions, but at the satisfaction of archival requirements for authenticity. This was a logical consequence of the premise that the technologies used in Cloud services are and will remain unknown. The only things we can be sure of knowing in a Cloud context is what is put into the Cloud for preservation and what comes out of it.

This kind of situation is one that has long been confronted in the physical sciences where there are many processes which cannot be observed directly. To deal with such processes, scientists use what is called a 'black-box' model. The process of interest is postulated as occurring within a black box, one that is opaque to any observation. All that can be observed are the inputs and outputs. What the process does, but not necessarily how it works, is inferred from the differences between inputs and outputs. The black-box model has also been adapted in computer science in the development of systems and applications. The adaptation addresses a process that does not exist, but must be created in order to satisfy a set of requirements. The requirements should specify the characteristics of inputs and the desired output(s). From the differences between the inputs and outputs, systems developers deduce what functionality and capacity the process must have. From this they can create or select hardware and software to satisfy the requirements.

The PaaST Approach

The PaaST project adapted the black-box model in another way. Given specification of what is to be preserved and the conditions of its preservation, the project asked what outputs are necessary to determine if preservation has been successful, and then articulated requirements that, if satisfied, would produce the specified outputs, without specifying either what technologies are used or how they work. The end product is a set of 1,350 requirements for the successful preservation of digital information.[3] These requirements are neutral with respect to the technologies that are used to satisfy them. The black box that contains the hardware and software can remain opaque. There is no need to know what products are used, how they are configured, or even when they are changed so long as the required outputs can be produced. In addition to being appropriate for the Cloud, or indeed to any situation where a contractor is unwilling to reveal details of its solution, technological neutrality maximizes the potential for effective automation; that is, it allows those who determine the solution, and how and when it changes over time, to choose the best technologies for the job.

Technological neutrality is also essential to address the root cause that makes digital preservation challenging: substantial, inevitable and unpredictable changes in information and communications technology. Changing

3 Kenneth Thibodeau, Daryll Prescott, Richard Pearce-Moses, Adam Jansen, Katherine Timms, Giovanni Marchetti, Luciana Duranti, Corinne Rogers, Larry Johnson, John R. Butler, Courtney Mumma, Vicki Lemieux, Sarah Romkey, Babak Hamidzadeh, Lois Evans, Joseph Tennis, Shyla Seller, Kristina McGuirk, Chloe Powell, Cathryn Crocker, Kelly Rovegno. Preservation as a Service for Trust (PaaST): Functional and Data Requirements for Digital Preservation. Version 1.0. November 11, 2017. https://interparestrust.org/assets/public/dissemination/PreservationasaServiceforTrust1_0.pdf

technology impacts digital preservation in two directions. First, technologies that work well to mitigate problems of digital preservation at present, or at any time in the future, are not exempt from obsolescence and will have to be replaced eventually. Second, advances in technology offer better ways to discover and use preserved objects. Preservation solutions need to incorporate the expectation of technological change both to overcome its negative effects and to benefit from the improvements it makes available.

The black box model that is necessary for using Cloud service providers might seem irrelevant when preservation is performed in-house. Any organization that takes on the performance of preservation activities knows what tools it uses. However, that is true only over a limited time span. No one can predict what technology will be like in two or more decades. Therefore, the further one looks into the future, the more an in-house solution darkens, eventually becoming a black box.

In addition to being technologically neutral, the PaaST requirements are designed to be adaptable to different situations, including different preservation objectives and policies, as well as various divisions of responsibility. For example, they allow for some aspects of digital preservation to be performed in-house, while others are contracted out, either on an ongoing or ad hoc basis.

The articulation of required outputs was a major challenge, given that the ITrust collaboration asked that the scope of PaaST be as broad as possible in terms of the kinds of digital information objects to be preserved and that it extend beyond the preservation of records; for example, to the preservation of data for scientific research in which data might be processed in ways that would be anathema in archival preservation. The project regarded this broad definition of scope as a desideratum, not an obligation. In order to support a broad scope, the requirements did not specify the preservation of records, but of "Preservation Targets," where a Preservation Target is, at least potentially, any digital information object whose preservation is desired. Nevertheless, it decided that to qualify as trustworthy, a preservation process should be able to output authentic copies of Preservation Targets. Starting from the archival definition of authenticity as the quality of an object "that it is what it purports to be and that it is free from tampering or corruption,"[4] PaaST converted this definition into terms that could be implemented in software: a digital reproduction is authentic if it is correctly identified and all of its properties that must must not change have remained the same from the time an object was ingested into a preservation environment. This formulation does not entail any assumptions or constraints on what a digital object is, why it is being preserved, what properties must remain unaltered, or how the identity and invariance are determined. While respecting the desired broad scope of the project, this formulation does not satisfy standards for articulating requirements for implementation in computer systems. It is simply too abstract and too broad. Put simply, it is not a requirement. Rather it is a fundamental criterion that guided the articulation of the PaaST requirements.

In this light, digital preservation is a process that enables the reproduction of a PreservationTarget and digital preservation is trustworthy if and only if it is capable of producing authentic copies of a Preservation Target. Determining that trust in digital preservation is merited requires empirical verification of the authenticity of

[4] "Authenticity" The InterPARES 2 Project Glossary. http://www.interpares.org/ip2/display_file.cfm?doc=ip2_glossary.pdf&CFID=21017256&CFTOKEN=81268255. Consulted 20200126.

reproductions. In the preservation of records in hard copy, there are different standards for the authenticity of copies applicable in different situations. The three principal standards are (1) copy in the form of the original, where the copy is essential indistinguishable form the original, (2) imitative copy, which retains the content and intrinsic and external form of the original but is explicitly a copy, and (3) simple copy, which reproduces the original in its entirety, but includes some changes, such as in font or page size.[5]

These standards can be adapted to, and are suitable for, application in the digital realm. They are suitable for application, in their established formulations, to Preservation Targets that are the digital equivalents of traditional documents, such as correspondence and reports. But they need to be adapted to deal with types of digital objects, and even properties of individual objects, that have no analog equivalent, especially functionality.

When the Preservation Target is software, it may be critically important to output copies that are identical to the original; that is, that have the same functionality, support the same types of user interaction, accept the same types of input and produce identical outputs from those inputs, and even have the same bugs as the original. Consider the case where a physician is accused of malpractice because of a misdiagnosis of a serious disease. The decision could hinge on the possibility of demonstrating that the physician made the correct diagnosis based on data output from a medical imaging system, but that the system's software included a bug that output incorrect or misleading data. Analogous concerns apply in other areas, such as when computer models are used in areas such as economic forecasts and government decisions.

In other situations, imitative copies might be entirely acceptable. Computer assisted manufacturing (CAM) systems, for example, are typically proprietary, incompatible with other systems and subject to obsolescence. Often the products of such systems, such as jet engines, ships and architectural elements, are maintained long after the CAM systems become obsolete. Maintenance of such products over time often requires piece parts to be replaced. If a newer CAM system is capable of imitating the obsolete system, in the sense of producing replacement parts that are identical to the originals, it does not matter if the newer system uses different hardware and software or if the system functions internally in a manner that is markedly different than the obsolete system it replaces.

The concept of simple copy is one that remains relevant, with only minor modification, to some Preservation Targets that do not have traditional counterparts. An obvious case is that of the worldwide web. The display of web sites varies depending on the devices used and even on individual users' settings on a particular device. Most websites today are designed to be accessible on different types of devices, ranging from smart phones through tablets to desktop computers. In this context, a copy in the form of the original is undefined. But simple copies, that include all of the content, organized as designed, preserving functionality, such as hyperlinks and user-determined parameters, are viable and appropriate.

In addition to the established concept of simple copy, however, it would be useful to introduce in the digital domain a related concept, that of simplified copy. All digital information objects depend to some extent on computer processing. At a minimum it is needed for creation, saving and retrieval. But one major element of this functionality should be excluded from digital preservation. The ability to create typically includes the capability to add, modify and delete. These capabilities should not be present in any copy reproduced from a

[5] Duranti, L., 1989. Diplomatics: New uses for an old science, Part I. *Archivaria*, *28*, pp.7-27.

preservation process. A simplified copy, then, is an authentic reproduction of a Preservation Target that cannot be altered in the state in which it is output from preservation.[6]

The appropriateness of different types of copies ultimately depends on the needs of those who request the copies. Institutions responsible for digital preservation should address the needs of their designated communities in determining what types of authentic reproductions are needed; however, practical considerations such as technical difficulty and costs may be decisive.[7] The PaaST requirements do not stipulate what types of authentic copies should be produced, but enable institutions to formulate preservation rules that enable automated implementation of their policies in this as well as other areas.

Verification of authenticity is a determination that the copy is correctly identified and that it has not changed in any significant way. Verification requires going outside the black box, specifically by capturing data about Preservation Targets and their preservation. Two categories of data constitute the foundation for verifying successful preservation: data that identify each Preservation Target and data that specify its properties that must not change. Both types of data must be captured starting at least from the time when something is designated as a Preservation Target. There are situations where those responsible for preservation do not have reliable data about Preservation Targets prior to their submission. In such cases, all that cane achieved and verified is successful preservation from the time the Preservation Targets are received.

The data needed to identify a Preservation Target vary depending on the objectives and policies that govern preservation. Identifying data can include both details about the Preservation Target itself, such as its genre, content, digital format, and internal organization, as well as data about the context or contexts in which the target was created or used. An important category of contextual data is that which differentiates Preservation Targets which, in themselves, are identical; for example, data identifying the different email accounts in which messages that were sent to multiple addresses are found. In the case of preserving records, in addition to data that characterize each record and differentiate it from other records, data about provenance and original order are needed. Understanding Preservation Targets preserved as sources of scientific information may require data that describe how research that generated a target was conducted, such as the research protocol, data collection methods, scope, frequency, precision and accuracy of observations; and details on how research data were transformed from initial observations to the forms in which they were designated for preservation, and how they relate to publications that disseminated research results.[8] Ensuring that a Preservation Target is adequately and correctly identified may entail preserving related objects. Preserving records requires preserving the record aggregates in which they were organized and kept by their creators. Preserving scientific data sets may require preserving research planning documents, traces of workflows used to process the data, and related publications.

---

[6] Of course, there is nothing to prevent anyone who receives a simplified copy from taking to another system where it could be altered, but that is outside the scope of preservation.

[7] Bettivia, R.S., 2016. The power of imaginary users: Designated communities in the OAIS reference model. *Proceedings of the Association for Information Science and Technology*, *53*(1), pp.1-9.

[8] Brinckman, A., Chard, K., Gaffney, N., Hategan, M., Jones, M.B., Kowalik, K., Kulasekaran, S., Ludäscher, B., Mecum, B.D., Nabrzyski, J. and Stodden, V., 2019. Computing environments for reproducibility: Capturing the "Whole Tale". *Future Generation Computer Systems*, *94*, pp.854-867.

Even within a single research project, in the case of multiple data files that conform to the same data model it may be necessary to preserve information that clarifies differences that are due not to the observed phenomena, but to factors such as different instruments, different locations where data were collected, or varying interpretations of data definitions in multiple collaborating research centers.

The Paast requirements, taken as a whole, define a comprehensive approach to verifiable preservation and, thus, establish a foundation for trust in digital preservation. The foundation comprises two parts: data about Preservation Targets and data about preservation. Data about Preservation Targets is collected from the start of the preservation process and used to verify the production of authentic copies. Data about preservation includes data about the processes executed for preservation and both data about the state of digital objects in storage and about whether preservation requirements and rules are properly implemented.

Figure 1, The PaaST Foundation for Trust, demonstrates how verifiably authentic digital preservation is achieved. Figure 1 assumes that data identifying a Preservation Target has been established by the start of the preservation process because that is necessary to establish basic control over the objects independently of preservation. The first step in building the foundation for trust is to specify what it is about a Preservation Target that must not change. These specifications are called Permanent Features in PaaST. The second step is to incorporate these specifications in the rules that determine when and how Preservation Services should be performed and then, thirdly, automatically implementing these rules the processes are executed. PaaST requirements further stipulate that when these processes are performed they generate data about what is done and which Preservation Targets are affected. The fourth step in building trust is using this data to assess whether the processes were performed properly and produced the appropriate results. The assessment is performed by Preservation Management. If it identifies any problems, they are used as feedback, in the fifth step, to correct problems in Preservation Services. If everything has gone as it should, it enables the sixth step, the production of authentic copies. The authenticity of a copy can be verified by comparing its attributes and operations with the Permanent Features of the Preservation Target. A reproduced PreservationTarget is authentic if all of its Permanent Features are identical to those specified when the PreservationTarget was designated for preservation.

Although figure 1 uses a digital image of a handwritten text as illustration, the PaaST requirements cover an unlimited variety of Preservation targets. The requirements divide potential Preservation Targets into two broad classes: machine-readable and human-readable. A machine-readable Preservation Target is one that is reproduced simply by loading it into a computer. In contrast, a human-readable Preservation Target must be loaded into a computer and then presented in a form accessible to humans. A human-readable Preservation Target is retained in the form of one or more machine-readable objects. For example, a textual document may be stored in a word processing file or a scanned image of text on paper; a table may be stored as a spreadsheet or a component of a database; however, many subclasses of machine-readable objects, such as software program, computer game, or digital model, have no analog equivalent. The most common factor that distinguishes suck classes is functionality that cannot be implemented in hard copy.

The specification of Permanent Features is essential and central to the PaaST approach to digital preservation. If you have not specified what properties must be preserved without change, you cannot determine if something

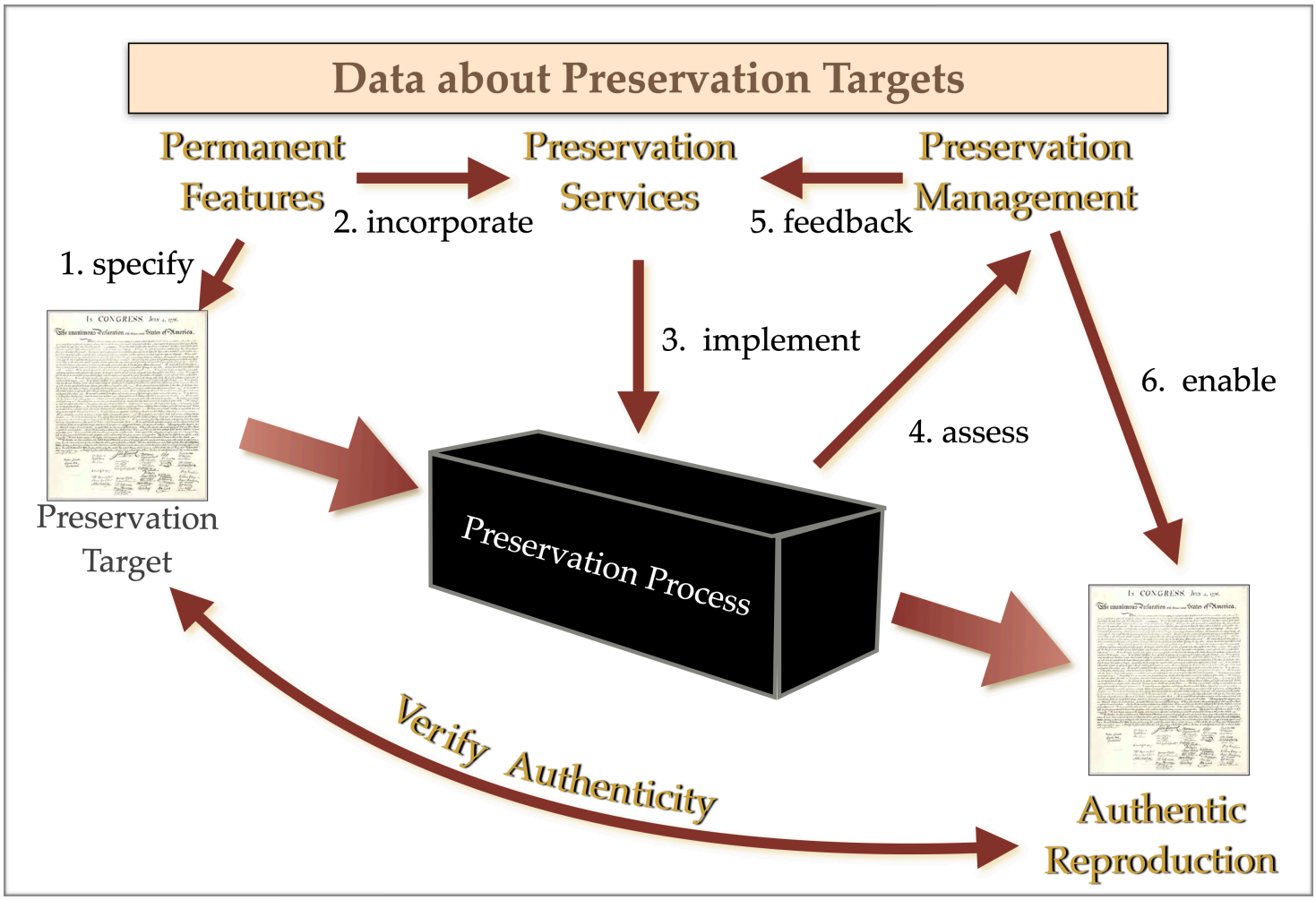

Figure 1, The PaaST Foundation for Trust

has been preserved. Permanent Features are either attributes or operations. An attribute is a static characteristic, such as content, structure or page layout. An operation is something that is done by or can be done to an object. Operations and attributes may be related. For example, a necessary operation for the preservation of digital video is playback, with a subsidiary operation of synchronization with audio. Any given video myst be played at a certain speed, which should be specified as a Permanent Attribute. Similarly, one or more attributes that define how audio is synchronized must be captured to enable successful playback. Even apparently static objects may have necessary operations. For example, a Preservation Target may be a quarterly report of activity of a certain sort. The presentation of the report in human readable form may depend on implementing a defined view on a relational database and applying the correct style sheet to generate the human-readable copy from the data output from the view. Implementing the view and applying the style sheet are operations.

To some extent, specifying Permanent Features is a matter of policy, grounded in professional judgment. An obvious example of this is determining which of the different types of authentic copies are required.

PaaST includes numerous requirements for specifying Permanent Features. Specifying Permanent Attributes involves three aspects: existence, value and expression, while specifying Permanent Operations involves defining their functions and any return value or postcondition that should exist when an operation has been performed.

Consider the example of email. Its Permanent Attributes include the identities of the sender, addressee(s), copy recipients and attachments. Do these attributes have to exist? All messages must identify the sender and at least one addressee. Otherwise, a message is incomplete. But the existence of other attributes is conditional: the identities of additional addressees, recipients of copies and attachments must be preserved only if they exist. What about the value of the Permanent Attributes? In the case of sent messages, the identity of the sender must be that of the account owner, but the identities of addressees cannot be specified a priori. Whatever identities are found in the "to" or "cc" fields must be preserved. The opposite is the case for received messages. The expression of a Permanent Attribute has two subfacets: how an attribute can be located within the stored digital object(s) that are needed to reproduce the Preservation Target and how they should appear when presented to a human.

The basic Permanent Operations for email, as for many types of human-readable Preservation Targets are retrieval and presentation. For the email of any user, the retrieval operation should be functionally specified as requiring the identification and loading of all messages in the user's account, and only those messages, the organization of the messages into folders established by the user, or default folders defined in the email application, the assignment of messages to the folders where the user put them, and the display of each message in appropriate format. Necessary return values include the correct sender, addressees, subjects and dates of each message and the inclusion of all attachments, if any. A required postcondition is the appropriate display of folders and messages.

Looking more generally, there are three types of universal Permanent Features that must be specified for all Preservation Targets. The first type relates to uniqueness; that is, the features that are inherent in a Preservation Target and distinguish it from any other and those that enable verification that the reproduction matches the data identifying Preservation Target. In the case of books, the distinguishing features would include title, author, date of publication, publisher, etc. In the case of software, the distinguishing features might be specifications of functionality that differed from one version of the software to another. The second type of universal Permanent Features are those that define the characteristics a Preservation Target should have when instantiated, either in a run-time version in the case of a machine-readable Preservation Target or format of presentation and interactive features in the case of a human-readable Preservation Target. For example, data preserved in a geographic information system should be displayed in map form and, if the original provided it, the user should be able to select the features that are displayed on maps. Similarly, a user should be able to rotate a three dimensional digital model through 360 degree solid angle. The third type of universal Permanent Features comprises those that relate to integrity, identifying all of the elements that are required to preserve and reproduce a Preservation Target, as well as the relationships among these elements. In the preservation of an archival fonds, for example, all the records that belong in the fonds must be preserved and it must be possible to assign each record to its appropriate place within the original order.

It might seem that the specification of Permanent Features is an onerous task, but it is an inevitable requirement for verifiable preservation. Moreover, to a large extent, Permanent Features can be specified for entire classes of Preservation Targets, such as archival fonds and scientific data sets preserved to support further research. Once specified, implementing the related PaaST requirements can automate much of the work required for Preservation Targets within those classes. Moreover, the feasibility and desirability of doing so is demonstrated

in the Victorian Electronic Records Strategy (VERS), which uses a similar approach entailing extensive data about the materials being preserved. VERS was implemented by the Public Records Office of Victoria, Australia beginning in 1998 and continues in use today.[9]

PaaST Requirements & Services

The discussion to this point focuses on the basic approach to digital preservation embodied in the PaaST requirements. What follows will shift attention to the requirements themselves. The number of requirements, 1,350, reflects adherence to basic standards for requirements engineering.[10] To provide a suitable basis for the development of computer systems and software, each requirement must stipulate one and only one thing; moreover, requirements collectively must be unambiguous. These standards are satisfied by organizing requirements hierarchically, with general rules refined with greater and greater specificity until all ambiguity is eliminated. Thus, PaaST defines a high level requirement to manage a Preservation Rule. But what does that mean? First of all, it must be possible to formulate a Preservation Rule that is suited to automated implementation. To clarify what is entailed in formulating Preservation Rules, PaaST articulates 27 related requirements that are decomposed down to five levels to ensure sufficient specificity and completeness. But these rules are irrelevant if they cannot be implemented once formulated. PaaST includes 28 more requirements specifying what is entailed in implementing Preservation Rules. They again are broken down to as many as 5 levels of increasing specificity. Similar situations exist for all 13 high level requirements, and in some cases the decomposition goes down to 7 levels.

One thing that distinguishes PaaST from other approaches to digital preservation is that the PaaST requirements do not define a preservation system, rather they define a set of related services. A service is a set of related actions that together are needed to accomplish a particular objective, function or task. PaaST defines 13 services in three groups: Preservation, Preservation Management and Information Management. Information Management services support both Preservation and Preservation Management, but the latter two can be implemented separately.

As Figure 2, PaaST Service Groups, shows, Information Management Services provide the basis for all other activities; Preservation Management Services are the framework for Preservation Services. Major elements of this framework are the definition and application of rules and reporting, assessment and verification.

There are four groups of requirements under Preservation Services: (1) Submission, which includes transfer of Preservation Targets, inspection to determine if they satisfy the applicable terms and conditions for transfer, and, if so, acceptance for preservation; (2) Storage, which includes the activities required for maintenance of the

---

[9] The VERS standard is available at https://prov.vic.gov.au/recordkeeping-government/vers.

Information on its development and implementation can be found in Waugh, A., Wilkinson, R., Hills, B. and Dell'Oro, J., 2000, June. Preserving digital information forever. In *Proceedings of the fifth ACM conference on Digital libraries* (pp. 175-184); Waugh, A., 2006. The design of the VERS encapsulated object experience with an archival information package. *International Journal on Digital Libraries*, *6*(2), pp.184-191; and Quenault, H., 2004, January. VERS: Building a digital record heritage. In *Archiving Conference* (Vol. 2004, No. 1, pp. 2-7). Society for Imaging Science and Technology.

[10] Institute of Electrical and Electronics Engineers. 29148-2018 - ISO/IEC/IEEE International Standard - Systems and software engineering -- Life cycle processes -- Requirements engineering. https://standards.ieee.org/standard/29148-2018.html

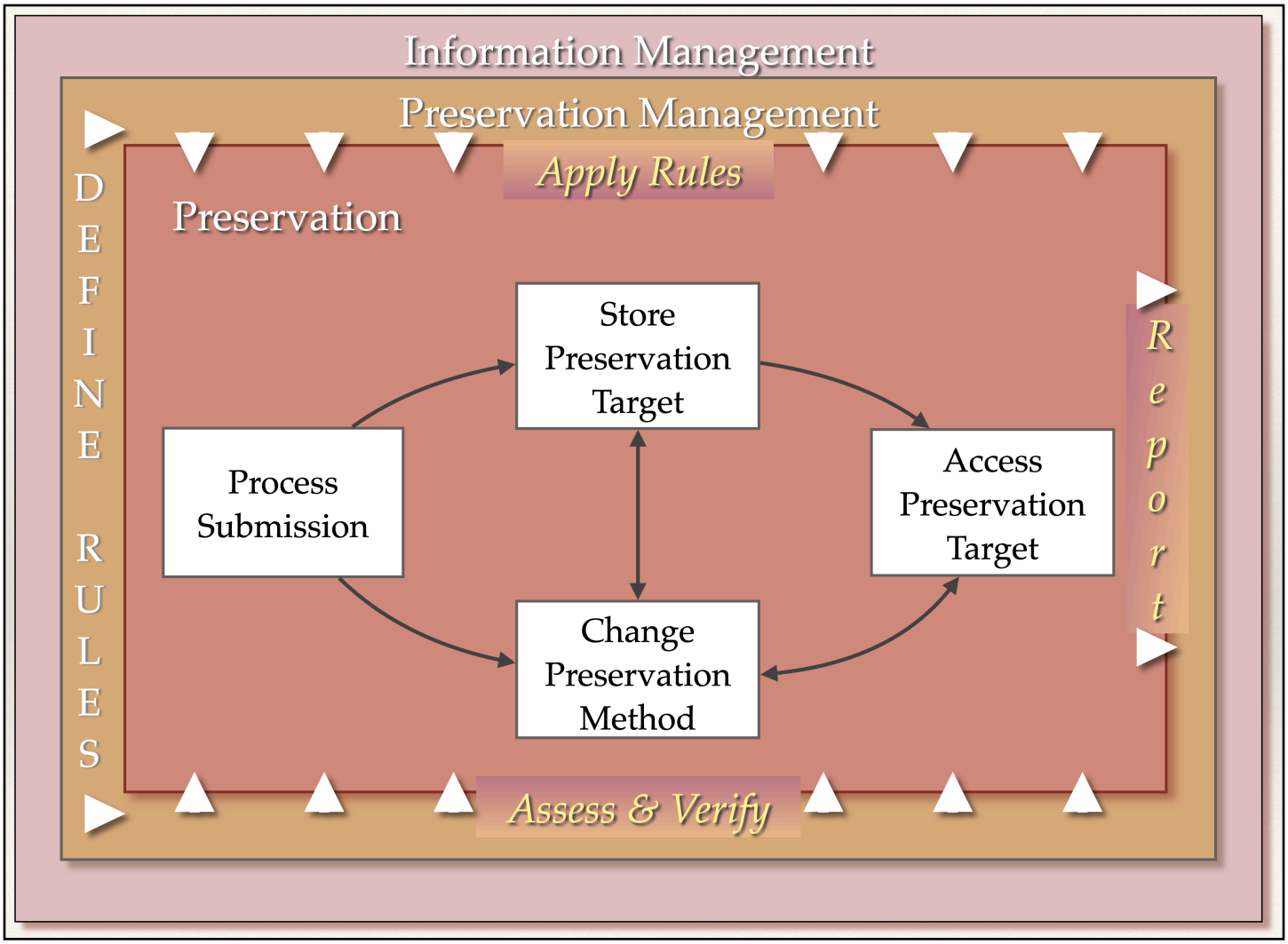

Figure 2. PaaST Service Groups

digital objects over time; (3) Access, which both enables authorized users to obtain information about Preservation Targets as well as reproductions of them, and implements restrictions on access; and (4) Preservation Change, which manages and implements responses to preservation problems and to hardware and software changes, including obsolescence.

Preservation Management Services also include four groups of services. (1) Preservation Rules which may cover anything and everything that can be done by other Preservation and Preservation Management services. This includes rules governing what should be transferred, when and how; who can request, authorize or perform what actions; and identification, enforcement & verification of Permanent Features. Assessment comprises actions for the inspection and evaluation of both the processes and the state of preservation. Assessment of processes covers actions such as determining whether processes are authorized and/or performed by qualified users, whether they are performed properly, and whether they produce the proper results. Assessment of the state of preservation includes actions such as determining if all digital objects that should be in Preserveration Storage are present, intact and properly identified; whether identifying data about Preservation Targets is

accurate and complete; whether data about Preservation Targets is appropriately matched to objects in storage; and whether current capabilities for reproducing Preservation Targets satisfy the requirements that derive from their Permanent Features. Verification encompasses actions that determine whether reproductions of Preservation Targets are authentic. Problem Handling requirements enable responses to problems identified in the performance of any other Services. They include capabilties for specifying when and how problems are identified and characterized; who may propose solutions; who must approve any such proposal before it is implemented, and what do do when a solution is implemented and either succeeds or fails.

Information Management Services include five sets of requirements. Three of them encompass generic capabilities, while the other two are tailored to specific needs of Preservation and Preservation Management. The generic capabilities are Data Management Services, Document Management Services and and Reporting Services. Data management, in PaaST as in general, enables and governs creation, access and updating of data; stores data and ensures their availability; and systematically implements policies for security, privacy, retention, disposition et al. Document Management Services support the definition of standard document forms needed in preservation including agreements with creators and donors of Preservation Targets, forms to be used in submitting materials for preservation, problem reports, preservation assessment reports and verification reports. PaaST Document Management also supports the production, sending and receipt of such documents, as well as insertion in and extraction of preservation management data from documents. Reporting Services enable the definition of required reports, their generation, transmission and review.

The two Information Management Services tailored for purposes of preservation are Class Management and Set Management. The difference between these two services derives from the different definitions of class and set in the PaaST data model. In PaaST a class is a group of things that have at least one, and often many, features in common. Email, for example, is a class in which every instance includes a sender, addressee(s) and transmission data and provision for a subject, message body and attachments. Similarly, all members of the class, relational database, share the organization of data into tables that are related by parent and child keys and they are all different from the class of graph oriented database. Classes may also be defined on the basis of features relevant to human use. Textual document, photograph and map, for example, are three different classes of document. Class Management Services enable both the definition of classes and the articulation of preservation rules and Permanent Features that apply to all members of the class.

In contrast to a class, in PaaST a set is a group of things that may have nothing more in common beyond that they are classified as belonging to a set. When they share other features, these often derive from membership in the set, rather than from features intrinsic to the objects themselves. Record aggregates, for example, often contain documents that belong to a variety of document classes, but they have attributes in common that derive from their status as records, such as records creators and the activities in which they participated, rather than from features of the class of documents to which they belong. PaaST distinguishes two different types of sets: Preservation Collections, which are sets that must be preserved as such, and Management Sets, which are defined for purposes of managing preservation. For example, if a given format becomes obsolete, a Management Set might be defined consisting of all digital objects in Preservation Storage that are in that format, regardless of what classes or Preservation Collections they are in, in order to perform a mass migration to another format. Also, a Management Set might be defined to include all Preservation Targets that are subject

to the same access restrictions. The PaaST requirements enable users to define, instantiate and manage sets that meet their needs. As with classes, rules and Permanent Features may be defined for sets.

PaaST Implementation

Some PaaST Services could be implemented in-house while others are performed on an on-going basis under contract. Still others might be performed only intermittently and by different contractors. When several parties have responsibility for different services, they do not have to use the same technologies, but can use solutions that are optimal for the services they perform. Commonality is required for functions like data management, document management, reporting and class and set management; however, while inputs and outputs must conform to the same specification, different technologies could be used to perform the required processing. Interoperability is required is some areas to achieve trustworthy preservation; notably if the formulation of Preservation Rules is done in a separate application, other services must be able to implement the rules and assess and verify their implementation.

The requirements allow broad flexibility in assigning responsibilities. Archives and other institutions that have the basic responsibility for preservation might prefer to formulate and manage Preservation Rules using in-house capabilities. Alternatively, capabilities for formulating and managing Preservation Rules could be included in a service contract, but the ability to exercise those capabilities restricted to users who are employees of the archives. In this arrangement, the service provider would implement the rules and report on their implementation. Digital storage might be provided under a long-term contract, while some actions, such as complex format conversions, might be assigned on an ad-hoc basis to other contractors with specialized capabilities..

In the case of institutions responsible for the preservation of records, intellectual control should remain in the purview of the archives since this function needs to be comprehensive and consistent for all holdings, not just digital ones. If one or more services are acquired by contract, the archives should retain sole responsibility for the establishment of Preservation Rules. Likewise, there are benefits, especially in terms of customer service, for an archives to have a comprehensive and coherent system for description of and access to its holdings. PaaST requirements enable the archives to establish data requirements so that data about records stored or processed by contractors can be integrated into such a system.

The service approach adopted in PaaST results in an extensive set of requirements covering all functions necessary for digital preservation, but provides wide latitude for archives to implement the requirements in accordance with local policies and norms. For example, the requirements support different specification of how archival concepts and norms such as ‘record,’ ‘record aggregate,’ ‘archival bond,’ ‘provenance,’ are defined in different institutions.

When services are acquired by contract, the archives should establish reporting requirements that enable it to determine conformance with its Preservation Rules as well as ensuring that contractual obligations are satisfied.