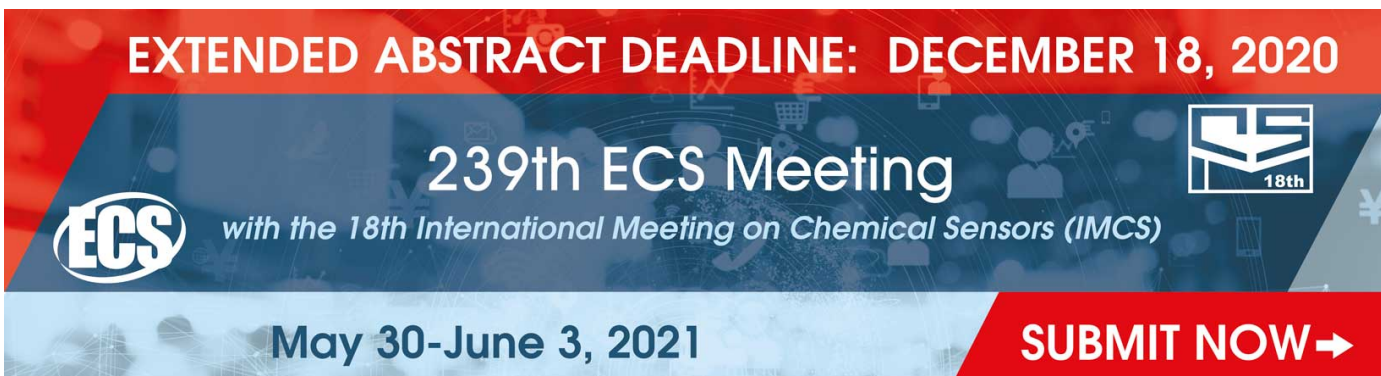


PAPER • OPEN ACCESS

Digital Preservation (challenges, preparedness and reaction)

To cite this article: F Perossini *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **949** 012112

View the [article online](#) for updates and enhancements.



EXTENDED ABSTRACT DEADLINE: DECEMBER 18, 2020

239th ECS Meeting
with the 18th International Meeting on Chemical Sensors (IMCS)

May 30-June 3, 2021

SUBMIT NOW →

Digital Preservation

(challenges, preparedness and reaction)

F Perossini¹, S Boi², MC Capua³,

¹CEO of Kpeople Research Foundation (MT), (perossini@kpeople.com
<https://orcid.org/0000-0001-9028-3179>)

²STORM Coordinator, Engineering Ingegneria Informatica SPA (silvia.boi@eng.it)

³Contracted professor ISCR (IT), (mariacconcetta.capua@gmail.com)

Abstract.

The STORM (H2020 700191), experience give us evidence on how it is important to address risks coming from natural hazards with the proper preparedness. Anyhow the project left one face of the picture uncovered. The aim of this paper is to stimulate further discussions related to digitised documentation and digital cultural assets safety either if they are digital born or reproduction of real ones. In 2003, UNESCO released the “Charter on the Preservation of Digital Heritage”[1], which defines digital heritage as “made up of computer-based materials of enduring value that should be kept for future generations (UNESCO 2003)”. Digital heritage emanates from different communities, industries, sectors and regions. Not all digital material is of enduring value, but a paramount volume of digital objects, especially the digital documentation related to cultural heritage, require active preservation approaches regardless the technology evolution. . It should be a priority to preserve that digital patrimony as well as to preserve the physical one. For this reason we aggregate the two view points in the “Digital heritage”. The challenge is represented by the preservation from one side of the digital born objects (including all the cultural heritage documentation), from the other objects coming from the conversion of existing/disappeared ones, must be protected and preserved. We are facing day by day a wider scenario where gallery, library, archive, and museum (GLAM)[2] dependence from digital content preservation is growing so the need for a lifecycle management of digital materials is necessary to ensure their continuity.

1. The discussion around Digital Preservation

The discussion on Digital Preservation has also an ethical aspect that deserves discussion: what should be preserved? – Can we consider the material of a CH physical object (i.e. marble) as the data format of a digital content? (see OAIS approach in the following part of the document) While addressing and comparing the digital heritage with the physical, historical and artistic cultural object, the discussion is complex: for the conservation of the cultural heritage basic principles followed by conservators refer to Cesare Brandi's hypotheses that does not separate the value of the work from its support and above all from its creative moment. The cultural object is closely connected with a time and a place that add a unique and absolutely unrepeatable connection; therefore any reproduction must be considered false. By translating these principles into digital preservation perspective, the format of the original data and its content represent, together with the time of the creative moment, the same unique and unrepeatable connection. Consequently, any kind of data conversion cannot be considered the original.

Since data conversion is necessary to guarantee future and present usability of historical digital data with current technologies, each content owner uses the methodologies he has selected to make it usable, without compromising the maintenance of the data in its original format.

2. Curation Lifecycle Model as the starting point



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

The Curation Lifecycle Model[3] has been developed as a generic, curation-specific, tool which can be used, in conjunction with relevant standards, to plan curation and preservation activities to different levels of granularity. STORM approach suggests to use as model the following steps: a) to be prepared, a preparedness tool for data creators, data curators and data users; b) to organise and plan adequate responses to natural disaster that could occur; c) to help digital heritage owners and managing authorities to identify risks to their digital assets and d) to plan management strategies for their successful curation. The figure on the right of Curation Lifecycle Model provides a

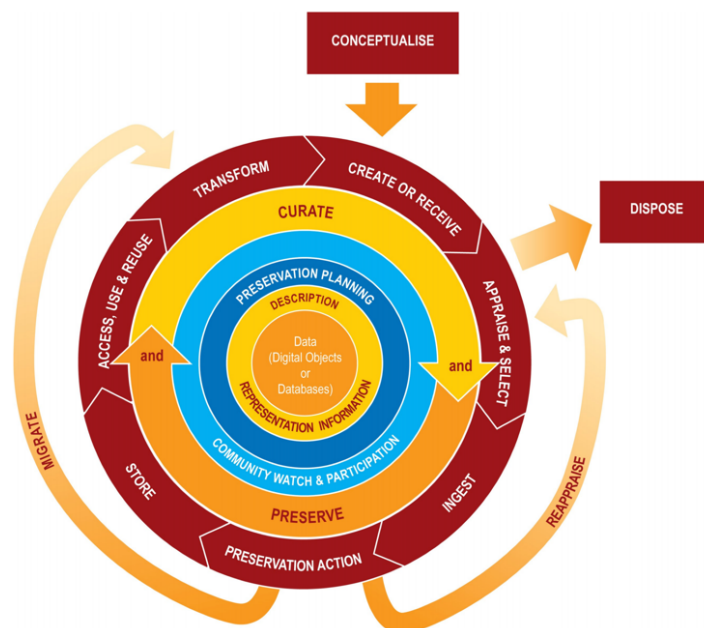


Figure 1. Curation Lifecycle Model

graphical high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt. The STORM team consider the Curation lifecycle can be used to plan activities within an organisation or consortium to ensure that all necessary stages are undertaken, each in the correct sequence able to provide also the proper preparedness for data curation responsible. Further research is needed to evaluate at what extent the model enables granular functionality to be mapped against it; to define roles and responsibilities, and build a framework of standards and technologies to implement. It can help with the process to identify additional steps which may be required to better address preparedness (certainly extensive training and STORM project risk assessment approach), or actions which are not required by certain situations or disciplines, and ensuring that processes and policies are adequately documented.

Curators should even consider that elements of the Curation Lifecycle Model are heterogeneous Data, (any information in binary digital form).

Curation lifecycle includes:

- **Digital Objects:** simple digital objects (discrete digital items such as text files, image files, videos, vectoral maps, 3D description, sound files, along with their related identifiers and metadata) or complex digital objects (discrete digital objects made by combining a number of other digital objects, such as websites, storytelling,).
- **Databases:** structured collections of records or data stored in a computer system.

FULL LIFECYCLE ACTIONS

Description and Representation Information	Assign administrative, descriptive, technical, structural and preservation metadata, using appropriate standards, to ensure adequate description and control over the long-term. Collect and assign representation information required to understand and render both the digital material and the associated metadata.
Preservation Planning	Plan for preservation throughout the curation lifecycle of digital material. This include plans for management and administration of all curation lifecycle actions.
Community Watch and Participation	Maintain a watch on appropriate community activities, and participate in the development of shared standards, tools and suitable software.

Curate and Preserve	Be aware of, and undertake management and administrative actions planned to promote curation and preservation throughout the curation lifecycle.
---------------------	--

SEQUENTIAL ACTIONS

Conceptualise	Conceive and plan the creation of data, including capture method and storage options.
Create or Receive	Create data including administrative, descriptive, structural and technical metadata. Preservation metadata may also be added at the time of creation. Receive data, in accordance with documented collecting policies, from data creators, other archives, repositories or data centres, and if required assign appropriate metadata.
Appraise and Select	Evaluate data and select for long-term curation and preservation. Adhere to documented guidance, policies or legal requirements.
Ingest	Transfer data to an archive, repository, data centre or other custodian. Adhere to documented guidance, policies or legal requirements.
Preservation Action	Undertake actions to ensure long-term preservation and retention of the authoritative nature of data. Preservation actions should ensure that data remains authentic, reliable and usable while maintaining its integrity. Actions include data cleaning, validation, assigning preservation metadata, assigning representation information and ensuring acceptable data structures or file formats.
Store	Store the data in a secure manner adhering to relevant standards.
Access, Use and Reuse	Ensure that data is accessible to both designated users and re-users, on a day-to-day basis. This may be in the form of publicly available published information. Robust access controls and authentication procedures may be applicable.
Transform	Create new data from the original, for example: - by migration into a different format, or - by creating a subset, by selection or query, to create newly derived results, perhaps for publication

OCCASIONAL ACTIONS

Dispose	Dispose of data, which has not been selected for long-term curation and preservation in accordance with documented policies, guidance or legal requirements. Typically, data may be transferred to another archive, repository, data centre or other custodian. In some instances, data is destroyed. The data's nature may, for legal reasons, necessitate secure destruction.
Reappraise	Return data which fails validation procedures for further appraisal and re-selection.
Migrate	Migrate data to a different format. This may be done to accord with the storage environment or to ensure the data's immunity from hardware or software

3. The importance of data classification

Another key factor to improve preparedness and offer the possibility of a clear estimation of possible damage suffered by data due to accident is to use standardised metadata; the General International Standard Archival Description (ISAD(G))[4] is a content model developed by ICA. ISAD(G) provides guidelines for creating descriptions of archival material, establishing a model based on the principle of respect within a multi-level description. ISAD(G) defines 26 elements in seven areas of a descriptive record (Identity Statement, Context, Content and Structure, Condition of Access and Use, Allied

Materials, Note, and Description Control), and provides general content guidelines. The ISAD(G) content model, along with the International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR(CPF)), serves as the foundation of Describing Archives: A Content Standard (DACS). So, a good classification of material is a mandatory starting point and it corresponds to what in the STORM project was the “description” phase in the Quick Assessment process. Here a sample of main classes of digital materials is provided:

1. Unpublished born-digital materials – personal and corporate papers, digital archives of significant individuals or institutions
2. Born-digital university archives
3. Research outputs – research data and publications (including compliance)
4. Published born-digital materials – physical format carriers (optical media), eBooks, web archives, archival and access copies of electronic subscription services, etc.
5. Digitised image materials – 2D photography (and 3D imaging)
6. Digital (and analogue) audio-visual materials – moving image (film and video) and sound recordings
7. In-house created content – photography and videography of events, lectures, photos of conservation treatments, etc.

The format of digital data is also another parameter that should be taken into consideration during the classification, here follow a non-exhaustive list of file format:

Main existing file formats[5]:

Still Image

- SVG_1_1
- TIFF_6

Moving Image

- MPEG-4_FF_2
- AVI

Textual

- PDF/A_family
- DOCX/OOXML_2012

Web Archive

- ARC_IA
- WARC

Sound

- WAVE
- MP3_FF

Datasets

- DBF
- HDF5

Geospatial

- ESRI_shape
- GeoPackage_1_0

Generic

- ASF
- RI

4. Digital Objects Self - Preservation as a AI challenge

Currently the level of automation in Digital Preservation solutions is low. The preservation process currently involves many manual stages, but should be approached in a flexible and distributed way, combining intelligent automated methods with human intervention. The scalability of existing preservation solutions has been poorly demonstrated; and solutions have often not been properly tested against diverse digital resources or in heterogeneous environments. These problems, together with the rapid obsolescence of software and hardware due to frequent update of private vendors, make Digital Preservation one of the most challenging application areas for Multi Agent System. In view of managing Digital Preservation, the prevailing paradigm consists of centralized data, and top-down approach where institutions are the main decision makers. To boost preparedness further efforts should be dedicated to propose a change of paradigm, mainly a bottom-up one, where the digital objects self-preserve: researching the possible use of Artificial Intelligence to support Digital Objects self-preservation, powered by a social network as a participatory environment enabling the attitude where “preservation is to share”. In this concept, DOs become active actors in their own Long Term Digital Preservation, here named the Self Preserving Digital Object (SPDO)[6], which has a DP budget devoted to funding the replication of the objects and other operations such as format migration or finding a safe storage within a social network of users; in all, an environment where they will “live”[7].

5. The OAIS reference model

In order to achieve a better advantage during the preparedness phase the selection of a consolidated reference model could represent a key success factor. For this reason OAIS has become “the reference model of choice of those involved in digital preservation worldwide,”[8] serving as a “galvanizing force”[9] and a “major factor in the advancement of digital archiving efforts.”[10] OAIS contribute to improve interdisciplinary sharing of information as “a common language and concepts for different professional groups involved in digital preservation and developing archiving systems.”[11]. In the meantime, the process shown in “The reference model represented common ground upon which to consolidate understanding of the needs and requirements of digital preservation: an opportunity to gather the strands of isolated digital preservation activities, merging them into a shared (albeit highly conceptual) characterization of the problem’s boundaries.”[12]

A typical information package includes the following information related to objects:

- Content Information: this includes the data object and its representation information
- Preservation Description Information: contains information necessary to preserve its affiliated content information (such as information about the item's provenance, unique identifiers, a Checksum or other authentication data, etc.)
- Packaging Information: holds the components of the information package together
- Descriptive Information: metadata about the object which allows the object to be located at a later time using the archive's search or retrieval functions

There are three types of information package in the OAIS reference model:

- Submission Information Package (SIP): which is the information sent from the producer to the archive
- Archival Information Package (AIP): which is the information stored by the archive
- Dissemination Information Package (DIP): which is the information sent to a user when requested

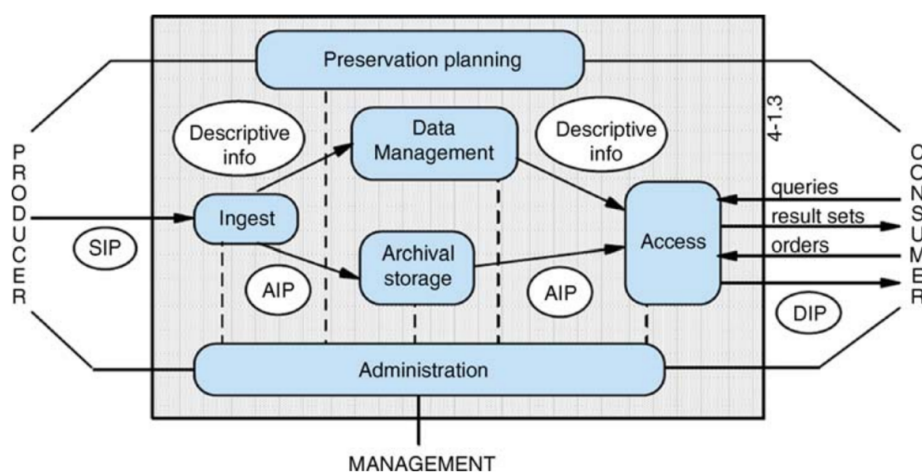


Figure 2. Information Package

OAIS Functional Entities. Source: Reference model for an Open Archival Information System (OAIS); Consultative Committee for Space Data System: Washington, D.C, 2002; 4:1. [SIP $\frac{1}{4}$ Submission Information Package; AIP $\frac{1}{4}$ Archival Information Package; DIP $\frac{1}{4}$ Dissemination Information Package]

6. Normalising metadata

Another important issue in running a digital preservation process is to normalise metadata to facilitate information retrieval, to support that need the ISO 15836-2:2019 Information and documentation —

The Dublin Core metadata element set — Part 2: DCMI Properties and classes, provides guidelines to address metadata issues. The Dublin Core Schema is a small set of vocabulary terms that can be used to describe digital resources (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like artworks[13]

- Title: name given to the resource
- Creator: entity primarily responsible for making the resource
- Subject: topic of the resource
- Description: account of the resource
- Publisher: entity responsible for making the resource available
- Contributor: entity responsible for making contributions to the resource
- Date: point or period of time associated with an event in the lifecycle of the resource
- Type: nature or genre of the resource
- Format: file format, physical medium, or dimensions of the resource
- Identifier: compact sequence of characters that establishes the identity of a resource, institution or person alone or in combination with other *elements* [SOURCE: ISO 8459:2009, 2.27]
- Source: related resource from which the described resource is derived
- Language: language of the resource
- Relation: related resource
- Coverage: spatial or temporal topic of the resource, spatial applicability of the resource, or jurisdiction under which the resource is relevant
- Rights: information about rights held in and over the resource

Conclusion

The overall conclusion is summarised in the following figure where four levels of preservation is presented. To be prepared for possible natural hazard affecting your site at least level 1 of preservation should be kept. Anyhow after the lesson learnt during the STORM project it is recommended to maintain level 2 or 3.

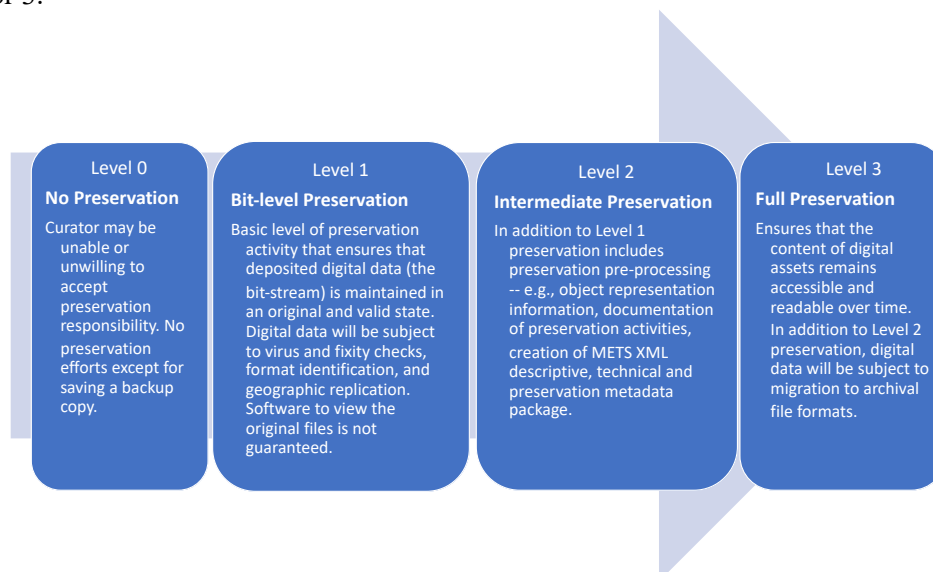


Figure 3. STORM levels.

Following principles assessed during the H2020 STORM project, exercises are suggested to verify the level of Digital Preservation adopted.

References

- [1] http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html
- [2] Langley, S. (2018). Digital Preservation Should Be More Holistic: A Digital Stewardship Approach. American Library Association.
- [3] Higgins, S. (2008). The DCC curation lifecycle model
- [4] <https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>
- [5] <https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>
- [6] de la Rosa J.L., Olvera, J.A.: First Studies on Self-Preserving Digital Objects. Artificial Intelligence Research & Dev. In: Procs 15th Intl Conf. of the Catalan Assoc. for Artificial Intelligence, CCIA 2012, Alacant, Spain, vol. 248, pp. 213–222 (2012)
- [7] Olvera, J. A., & de la Rosa, J. L. (2015, June). Addressing long-term digital preservation through computational intelligence. In International Conference on Practical Applications of Agents and Multi-Agent Systems (pp. 304–307). Springer, Cham.
- [8] Greenstein, D.; Smith, A. Digital preservation in the United States: Survey of current research, practice, and common understandings. In *New-Model Scholarship: How Will It Survive?*; Smith, A., Ed.; Council on Library and Information Resources: Washington, DC, 2003; 40–48, 43
- [9] Waters, D. Good archives make good scholars: Reflections on recent steps toward the archiving of digital information. In *The State of Digital Preservation: An International Perspective*; Council on Library and Information Resources: Washington, DC, 2002; 78–95, 80
- [10] Hodge, G.M. Digital Preservation: Overview of Current Developments. *Inform. Serv. Use* 2002, 22 (2/3), 73–82
- [11] Beagrie, N. National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity; Council on Library and Information Resources and the Library of Congress: Washington, DC, 2003, 45
- [12] Lavoie, B.F. The Open Archival Information System Reference Model: Introductory Guide; OCLC Online Computer Library Center and Digital Preservation Coalition: Dublin, OH, 2004; 2.
- [13] <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-5>