

*Qualitative and Quantitative Methods in Libraries (QQML) 5: 695- 710, 2016*

---

## Hold it All Together: a Case Study in Quality Control for Born-Digital Archiving

Paul Gabriele Weston, Emmanuela Carbé and Primo Baldini

University of Pavia, Italy

**Abstract.** PAD (*Pavia Archivi Digitali*) is a project developed at the University of Pavia aiming at long-term preservation of born-digital private and literary papers produced by Italian writers and journalists. Since 2009, when the first collection was entrusted to PAD, the archive has grown and it currently includes almost 80,000 files. Dealing with this material is inherently complex, as it implies the integration of fields such as literary studies, archival techniques, information technology, and legal and administrative matters. In particular, evolution in technology, use of different types and versions of software, ever changing storage techniques, from floppy disks to cloud services, are all factors posing severe challenges. To ensure a rigorous management of the workflow, which is critical to the success of the entire project, a specific piece of software - QUANDO (*Quality control for Archiving and Networking Digital Objects*) - was developed.

**Keywords.** Private Papers Archiving, Quality Management System, Quality Control Software, Born-Digital Archiving, Digital Curation, Evaluation, Qualitative Methods

### 1. Introduction

Collecting and archiving digital-born papers is by no means an easy task. Evolution in technology, use of different types and versions of software, ever changing storage techniques, from floppy disks to cloud services, all pose a series of severe challenges in securing preservation for collective and personal memories of the last decades. In fact, a project aiming at long-time preservation of such material is inherently complex as it requires the integration of competencies belonging to areas that are considerably different from one another: literary studies, archival techniques, information technology, and legal and administrative considerations. Furthermore, its management implies an endless upgrading of data models, standards and procedures to cope with the growing variety of data sources. In defining “born digital” and the various types of born-digital materials, Erway (2010) neatly describes the scenarios and the issues encountered in preserving files produced by combining such a variety of formats. As compared with challenges usually associated with the preservation

Received: 10.7.2016 Accepted: 21.3.2016

ISSN 2241-1925

© ISAST



of traditional manuscripts, handling born-digital materials includes risks such as bit rot, obsolescence of media, hardware and software, and loss of integrity of data. These technical issues are enhanced by factors of a more administrative nature concerning levels of responsibility, legal matters, and finally accessibility by users. Pavia Archivi Digitali (PAD) is currently seeking solutions that will help guarantee the feasibility of the archiving project and, at the same time, get on top of these problems.

## **2. The Pavia Archivi Digitali project**

The Pavia Archivi Digitali project, also known under the acronym PAD, was developed at the University of Pavia in view of the long-term preservation of the born-digital literary and personal archives produced by a certain number of Italian writers and journalists. The structure currently falls under the responsibility of the University library system, but it is quite reasonable that in the years to come it will become part of the Centre for Research in the Manuscript Tradition of Modern and Contemporary Authors (Centro Manoscritti, Pavia University). This research centre, which has quite a reputation in Italy among scholars in the humanities because of the number of archival fonds it is committed to preserve and put at the disposal of research and teaching communities, was originally created by Maria Corti, who besides being a philologist was a writer in her own right. In 1969 Maria Corti got the idea, which was quite innovative at the time, of collecting the working papers of contemporary Italian fellow poets and writers. Years later, in 1997, she wrote a book, *Ombre dal fondo*, in which she described how her initiative got started.

The origin of PAD dates back to year 2009. It was the journalist and writer Beppe Severgnini who proposed the creation of an archival system for the preservation of the otherwise undisclosed born-digital textual material provided by Italian authors. In his vision, PAD should be set up to host a wealth of memories that, whilst providing scholars with sources for investigating Italian literature and culture of our time, could progressively be made accessible to the community at large. The reason for creating a new structure, separate from the existing Centre, is to be traced in the specific features and needs of digital files, quite different from the paper-based archives that the Centre has treated so far. PAD should therefore be given an organization and enough flexibility to address in an appropriate, more efficient way the new issues, which of course require a specific approach. In 2011 Severgnini bestowed to PAD more than 14,000 files downloaded from his personal computer. Over the years, more authors have transferred to PAD their literary archives (among them Silvia Avallone, Gianrico Carofiglio, Paolo Di Paolo, Francesco Pecoraro and Franco Buffoni). Almost 80,000 files are currently preserved in PAD.

## **3. The core issues**

Despite the enormous growth in digital content being created by individuals, the amount of research and development dedicated to the preservation of this kind of material is currently visibly insufficient. Large digital preservation projects

either sponsored or directly managed by corporate bodies or state archives have led to the development of policies, standards and good practices, procedures, and archival repositories designed to preserve digital objects deemed to be of great value to the organisation itself. Preserving Access to Digital Information (PADI), maintained by the National Library of Australia, provides us with a good example of these efforts. PADI, whose activities were closed by the Advisory Board in August 2011, was in fact a gateway to digital preservation resources and at the same time acted as a safekeeping device by bringing resources deemed to be of long-term value together within the PANDORA web-archive, where they will supposedly remain fully accessible. In carrying out this huge work, organizations such as the National Library of Australia were moved by legal, financial and organisational drivers which in general play little or no role in the preservation of personal papers.

Working with individuals to preserve their personal digital archives presents a variety of challenges which require *ad hoc* solutions. In the first place, personal papers are not just born-digital manuscripts. Quite often the pre-textual phase implies the production of a mass of documents which include photographs, cartographic material, recorded interviews, scrapbooks and so on. Not all of this content is immediately transferred onto the written page. Some writers store them properly in folders in case they prove useful for later use. On the contrary, if they are scattered throughout the archive, the logical link with the original work is often lost. Besides, a literary archive is also a place where unfinished works are saved only to be reopened from time to time simply to add a name, a character, a chapter, in which case a manuscript is generated through the superimposition of layers of scripture, each of which should be preserved and documented. Unbundling a manuscript having this nature can be incredibly troublesome. Ideally, the digital curator should work with donors to prevent this problem to occur (Weston 2013, p. 56-65). The writer should be advised in advance on the most efficient ways of weeding, organizing, and naming files, on recommended formats and media. He or she should even seek to preserve the equipment, as well as the media. Strictly speaking, digital curation “involves the management of digital objects over their entire lifecycle, ranging from pre-creation activities wherein systems are designed, and file formats and other data creation standards are established, through ongoing capture of evolving contextual information for digital assets housed in archival repositories” (Lee and Tibbo 2007. See also Lee, Tibbo and Schaefer 2007, 49-50). This definition appears to be too restrictive in the case of a literary archive, provided that no archivist would ever dream to dictate to an author the way in which he or she should write his or her book. Are we not allowed to use the term “digital curation” then? Cragin (2007) has provided a definition which seems much more adequate to describe the activities run internally at PAD in fulfilling its mission: “active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education; ... activities [which] enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time”.

#### **4. Preserving the context**

One of the most critical issues in handling digital files for long-term preservation purposes lies in the preservation, at the most detailed level of granularity, of the environment where the archives were generated and hosted. Because of the porosity of our digital world, there is little demarcation between storage methods, delivery mechanisms, and the devices with which we access, read, and interpret our sources. There is a strong need for better understanding causes and consequences of these phenomena if new, more adequate policies should be introduced to safeguard our cultural heritage, especially considering that about 90 percent of our records today are born digital. Given the ever growing amount of digital-based documentation in the humanities and digitally produced and versioned primary sources, interpreting, preserving, tracing, and authenticating these sources requires the greatest degree of sophistication. Should we then turn to digital forensics to get some hints? When the Digital Forensic Research Workshop met in 2001 the Digital Forensic Science was defined as “the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal.” As anyone can appreciate, the line separating the kind of information subject to forensic analysis and that of interest for literary or historical studies is very thin. Moreover, data, machines, and methods appear to be agnostic, which adds one more reason in favour of this interdisciplinary approach. As Kirschenbaum, Ovenden and Redwein (2010, p. 1) put it “the methods and tools developed by forensics experts represent a novel approach to key issues and challenges in the archives and curatorial community.”

PAD, as any archive of a similar nature, is receiving computer storage media (and sometimes entire computers) as part of its acquisition of “papers” from contemporary writers, journalists and scholars. Smartphones, e-book readers, and other data-rich devices will follow before not very long. A piece of software emulating the one used to index the contents of a criminal suspect’s hard drive is being tested to prepare a comprehensive picture of the electronic files donors have bequeathed to PAD. In a similar way, the hardware that allows the forensics investigator to create an algorithmically authenticated “image” of a file system allows the archivist at PAD to ensure the integrity of digital content once downloaded from its source medium. The reconstruction of a lost or inadvertently deleted version of an electronic manuscript can be carried out by using the same data-recovery procedures that allow the specialist to discover, recover, and present as trial evidence an erased file. Despite the fact that by combining all these techniques we are provided with a trustworthy and precise insight into the author’s literary archive, to obtain from the start a better picture of the structure of the files and to understand how the author’s writing habits and his or her communicational strategies have influenced the setting of the archive, PAD staff have prepared a questionnaire which is submitted to the

donor prior to onsite evaluation and appraisal and to transfer of files for archival custody. In this phase, the assistance from the author in person is extremely useful.

Having prepared a biography and a bibliography, these are submitted for consideration to the author who is asked to fill in an informational survey. Despite PAD cannot handle the information provided as ultimately exact, the survey is a useful tool and a good starting point. The survey covers 15 main areas: (1) shape of the archive to be deposited in PAD; (2) differences with respect to the author's personal archive; (3) structure and contents of the bestowed archive; (4) terms of consultation; (5) terms of privacy (the author is also asked if he or she recognizes third-party generated files or e-mails received); (6) incrementability of the archive with further donations; (7) further bestowed objects (such as web pages, social network contents, newsletters); (8) information about the computer in use; (9) information about previous devices; (10) information about operating system in use or used in the past; (11) other devices and secondary computers in use; (12) organization of the archive (whether he or she knows how documents are organised, whether files are saved to particular locations during the activity); (13) word-processing software mostly used and how he or she works with their editors; (14) activity in the web; (15) use of social networks.

When the archivist at PAD gets to the bestowal location, which in most cases coincides with the author's residence, all technical information regarding the storage devices and the electronic equipment, which had been provided by the author him- or herself through the survey, is checked and validated in the first place.

Based on the type and the size of the archive to be transferred, the archivist has to make some decisions about the preservation of the original context at the time of the bestowal. When the author's private files are tightly combined with the literary files, preserving the context can be somehow problematic and could require some drastic decisions on how far to push the context safeguard action. In what case is a formal description of the private files absolutely necessary to give sense to the rest of the archive? how do we make sure that no useful content is deeply embedded in the private part of the repository? In fact, the ideal situation is when the private and the literary archives coexist within the same computer, but are clearly distinct. Francesco Pecoraro's case is emblematic, as he transferred almost fifty thousand files. In order to preserve and to be able to rebuild the original context, the archivist had to take screenshots of the desktop and of the applications showing the file system structure, and had to take note of the folders that the author had not allowed to transfer for personal reasons, along with some information about what he was not bestowing.

In other cases (namely Gianrico Carofiglio and Silvia Avallone) the number of files involved in the donation was considered too small to require contextual information.

During the bestowal process the archivist takes notes of every single operation, the exact time when they are performed, the involved devices, as well as of technical or administrative problems that may occur (Erway 2012, p. 3-4). Pictures and audio-videorecordings of the event are also taken for PAD record.

The QUANDO software, which is described in the following chapter, easily enables reconstruction of the preservation history of a fond, from the moment it was taken in custody to when it was downloaded to PAD facilities in Pavia. All communications between PAD and the authors (e-mails and telephone calls) are recorded as well, and so are as the decisions made by the scientific committee about the issues concerning the management of the project in its various aspects (Fig. 1). The entire archival history is entered and preserved in the Info area which is linked to the relevant fond.

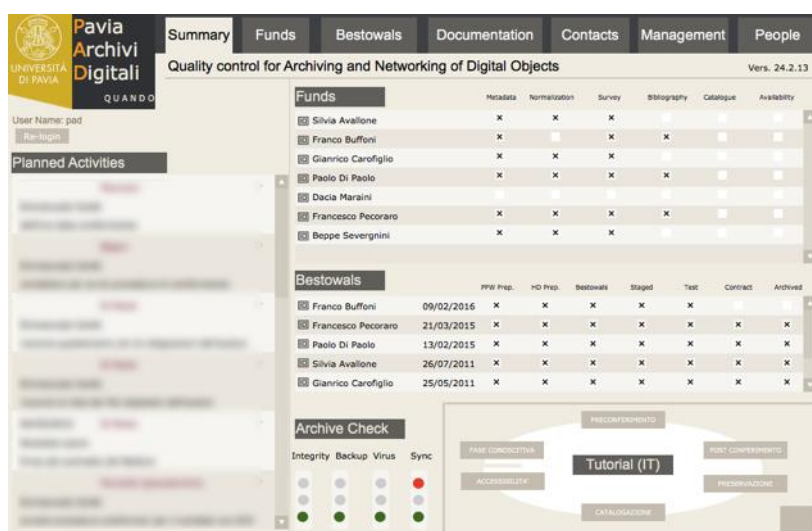
### **5. How to “hold it all together”? The “QUANDO” software**

Given the substantial investment in time and expertise required by the setting of PAD archival system, it is crucial that all institutional activities regarding data transfer, analysis and archiving, media production and digitization be identifiable and traceable in the most accurate way. Within PAD, digital contents include texts, audio-visual recordings, web sites and blogs, metadata, pieces of software designed to carry out specific procedures, rights management instructions, user profiles. As a whole they represent PAD’s digital asset. Since the volume of digital files transferred, created and maintained by PAD is dramatically growing, their management becomes increasingly important if staff are to easily retrieve and use them. Digital asset management procedures include activities associated with the creation, cataloguing, storing, retrieving and backing up of these assets. Their purpose is to integrate best practices within workflows to improve access to resources and make them available for reuse. One of the basic functions carried out through a digital management system concerns file-naming procedures and storage conventions, so that files can be easily tracked, located and retrieved for different purposes. The coexistence of several versions of the same work generated by the cumulation of superimposed layers of the author’s archive could cause great frustration when trying to identify and locate the individual files in chronological order of creation or transfer. A management system should help to avoid these scenarios by establishing consistent file management practices within the organization. File management is but one of the areas whose efficiency can be improved by the adoption of digital management assistance. Other areas include metadata management, workflow, policy tracking and enforcement, and access (van Niekerk, 2006).

A management system of this kind involves the creation of an archive, the development of an infrastructure to preserve and manage digital assets, and search functionalities allowing end users to identify, locate and retrieve an asset.

At its simplest, this system consists of a set of database records. Each database record contains metadata providing the name of the file, its format and information about its content and usage. Digital asset management software can be used to create and manage the database and help the institution to store simple documents as well as rich media in an affordable manner (Magan 2005).

**Fig. 1: QUANDO. Summary display**



When time came for PAD to adopt a DAM software, the acquisition of a commercial solution was turned down as this was far beyond available resources and because these solutions are usually tailored on the needs of a business company. We were also quite reluctant to take on software providers claim to be “free”, requiring nothing more than a “double-click” installation. DAM software has to be integrated into PAD’s processes, meaning that even after the free download and the double-click installation, configurations must be made, and adjustments will be required over time to fine-tune the system to keep pace with the archive’s evolution. As Diamond (2012, p. 7) remarks: “The fact is, when done right, DAM can help organizations save time and money, better leverage their investments in digital assets, and even avoid litigation. And when done wrong, DAM can cause production delays, destroy creative workflows, frustrate users [...]. DAM is so deeply integrated into core business processes that most experts agree the best way to increase your chances of digital asset management success is to wrap DAM into a corporate initiative you design, manage, deploy and promote exactly as you would if you were a software developer releasing a new product”.

The creation of a locally designed DAM software began with the development of a Quality Control Software, dubbed QUANDO (Quality control for

Archiving and Networking Digital Objects), aiming at supporting procedure supervision in the first place. QUANDO, based on a FileMaker platform, was initially implemented as a stand-alone application and subsequently ported to a multi-user environment accessible via the Intranet. It is mainly developed for personal computers and iPad tablets, but can be accessed through almost any device. The software is able to manage all aspects of the process, integrating information entered manually with data that is gathered automatically using a PAD-developed application suite performing a variety of actions on every single archive: checksumming, antivirus scanning, metadata extraction, synchronization, etc.

Over time QUANDO was constantly enhanced to become a proper DAM system. Besides data treatment, QUANDO supports, in fact, the many actions that the management of the literary archive implies: acquisition of the files, convalidation and authentication, preservation management, coordination of staff (academic board, PAD administration, repository management, archivists and cataloguers, legal consultants); ongoing testing and improvement of the entire workflow.

## **6. The repository architecture**

The System Architecture for the preservation and handling of files and data has been designed in compliance with the recommendations of the Open Archival Information System (OAIS) reference model. As Lavoie (2014, p. 22) reminds us: «OAIS is a model and not an implementation. The standard says nothing about system architectures, storage or processing technologies, database design, computing platforms, or any of the myriad technical details involved in setting up a functioning archival system. However, since OAIS publication, a number of initiatives have used it as a building block – a conceptual foundation and starting point – in support of the construction of functioning archiving systems».

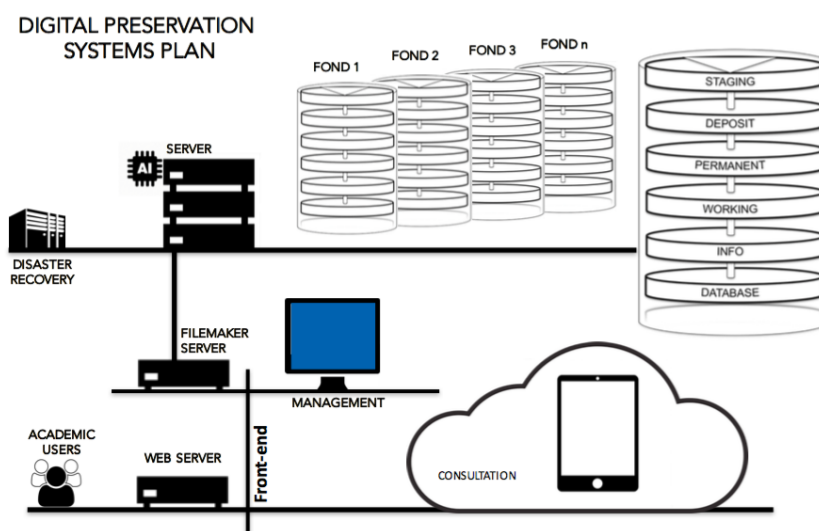
In PAD's perspective, OAIS conformance should imply much more than the application of OAIS concepts, terminology, and the functional and information models in the course of developing the digital repository's system architecture and data model. Relying upon the efforts of the Task Force on Digital Repository Certification, which was charged with developing criteria to identify digital repositories capable of reliably storing, migrating, and making available the portion of the scholarly and cultural record in their custody (TRAC 2007), PAD adopted TRAC checklist as revised and published as CCSDS 652.0-M-1 (Magenta Book) and subsequently approved as an ISO Standard (ISO 16363) (CCSDS 2012). This checklist, whose scope covers the areas of organization and governance, management of digital objects, and technology, represents best current practice and thought about the organizational and technical infrastructure required to be considered trustworthy and capable of certification. A set of requirements was established in the first place, taking into account the growing needs of the archive and the wide spectrum of functionalities the implementation of which one might eventually consider desirable: 1. a simple



and scalable architecture; 2. fast archival of big masses of data; 3. archival procedures carried out also in case of missing identification metadata; 4. document conversion in a non-interactive mode; 5. automated metadata extraction; 6. restricted access to contents whenever specified in the agreements with the author; 7. display of the filesystem structure of donations; 8. handling of redundancies, exceptions, occurrences, layering, etc.

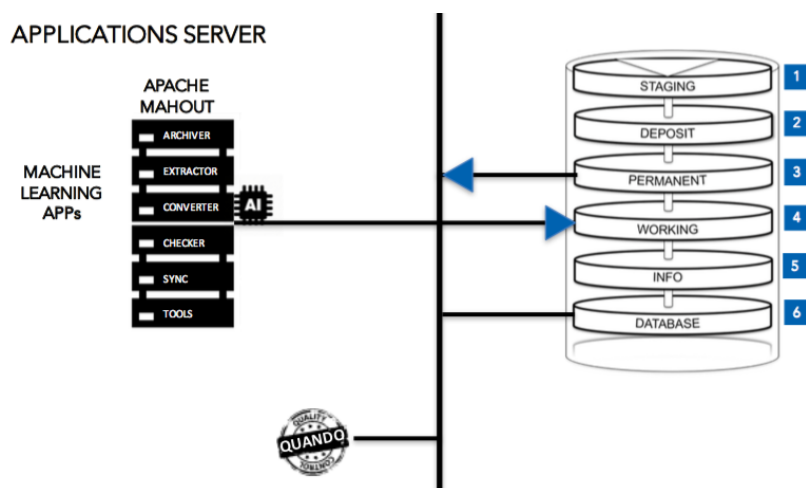
Based on these requirement, PAD management system was developed on an Enterprise Content Archive platform and entirely written in Java, which makes it easily customizable, flexible and extensible (Fig. 2). Each fond is archived, managed and displayed as a separate entity, enabling the implementation of distinct access policies. The archival system is based on five areas: Staging, Deposit, Permanent, Working, Info. Having been processed, each fond is allocated a separate database.

**Fig. 2: The archival system structure**



A set of applications (PADCleaner, PADList, PADArchive, PADPrint, PADChecker, PADEXtractor, PADConverter, PADSyc) reside on the java server. Each of these apps is responsible for a specific function. All procedures are constantly monitored through the Quality Control software QUANDO and are subject to the access policies established either at the fond or at the individual file level (Fig. 3).

**Fig. 3: The main operations are located in the “read-only” area (named “Permanent”) and in the “Working” area.**



The first two areas are temporary storage areas. Upon arrival, files downloaded directly from a hard drive or from cloud storage, or coming from social media, mail servers, and the web, are preserved in the staging area while waiting for an operator to be available (Fig. 4). Under the control of the latter, the files are then moved to the deposit area, where the archive integrity is checked, existing viruses are isolated, and hidden files are removed (unless they contain information that could be extracted and used for metadata creation). The PADPrint app generates a list of files, that once printed is sent back to the author for his or her consideration. On the basis of such list, the author can decide which files should not be considered part of the literary archive and therefore removed or be subject to embargo (Fig. 5).

Having carried out all these procedures and purged the files as instructed by the author, the archive is copied to the permanent area and it undergoes an initial process finalised to its preservation: an unencrypted copy of the archive is burned onto Gold Preservation quality DVDs and transferred to a vault located in a bank. In addition, for every archive one copy is stored in the main site of PAD at Pavia and another one at the university site in Cremona, more than fifty miles away, thus following Distributed Digital Preservation good practices. Only at this stage, is data residing in the two temporary areas definitely removed.

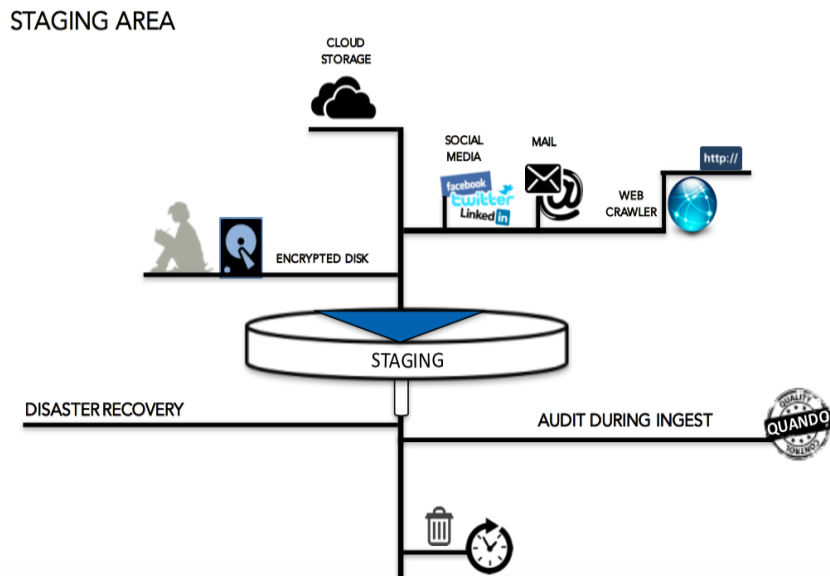
Files are then copied to the Work Area where several procedures take place: technical metadata are extracted and documents are converted to formats allowing for longer accessibility (Fig. 6). Whenever possible and opportune,

older computers and configurations are emulated using virtualization technology (Fig. 7). The strategy involves the file conversion into three different formats: Open Document Format (ODF) - Standard: ISO/IEC 26300:2006 (chosen as the most “open” format available); Portable Document Format (PDF) - Standard: ISO 32000-1 (in reason of its widespread use and readability); Text file (TXT)-Standard: ISO-8859 (to provide file statistics, as well as optimising the preview function). Converted data, together with metadata, both saved in XML format, are stored in the Permanent Area alongside the original files.

The Permanent Area is constantly virus checked and checksum tested to ensure the integrity of the archival fonds.

Eventually, the entire documentation related to the bestowal procedures or collected by the QUANDO system is transferred to the Info Area, including videos, snapshots and other multimedia files, as well as data generated in the process and not belonging to the author’s own archive.

**Fig. 4: The intermediate storage area between the sources of information and the repository. It is usually of a temporary nature, and its contents can be erased after the repository has been successfully loaded.**

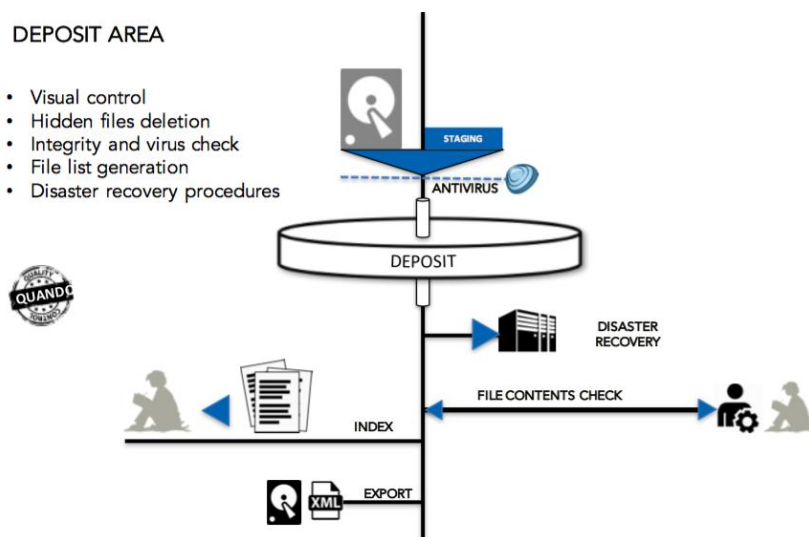


PAD is currently working on the PADExplorer application, a tool aimed at accessing and describing documents. Among its features is the possibility of displaying data from different sources (i.e., the five areas and the database), of applying filters, and including further options -- such as occurrence lists, metadata visualization and PDF preview.

### 7. Open challenges and new perspectives

The PAD project includes several groundbreaking features which make its development quite a challenging task in many ways. Description of fonds in line with the archival tradition, digital treatment of data, methods and conditions of use for research and study, preparation of learning objects targeted to schools, legal matters, long term preservation policies are issues that deserve a lot of consideration, so they cannot be all addressed simultaneously. On the other hand, PAD is not provided with all the resources and the skills needed. To foster a wider knowledge of its existence among specialists, and to seek support from experts in the various areas, PAD has organised technical workshops each focused on a specific matter. Which criteria to adopt in archiving both files and contents in the long term preservation perspective was the topic of the first event. The following workshop was dedicated to legal matters, an issue which the hybrid nature of data and the frequent case of joint responsibilities involving the existence of multiple rights owners make particularly tricky.

*Fig. 5: Procedures carried out in the Deposit Area before files can be written to the Permanent Area*

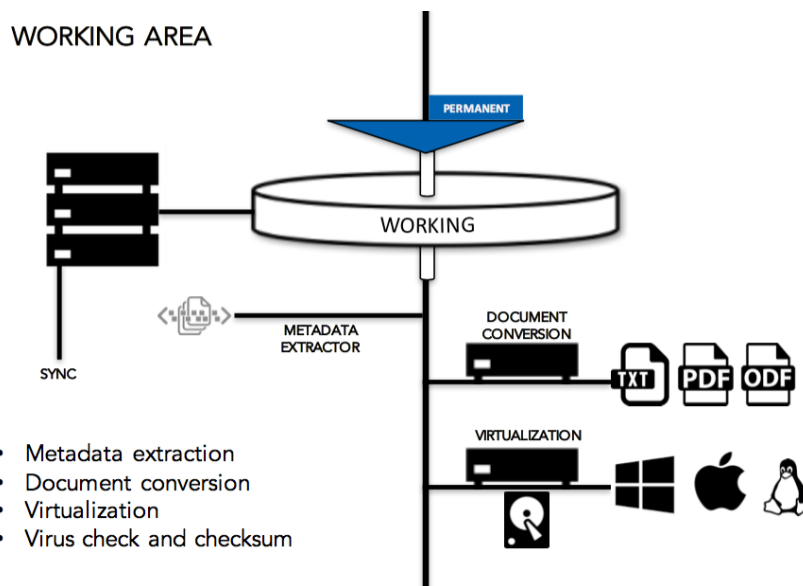


In parallel, literary evenings targeted to a wider audience are organised and the authors who have bestowed their papers to PAD are invited. They can then exchange views with journalists, publishers, students, ordinary people about the role that digital media play in their professional and social lives, and how the use of pcs' and mobile devices has changed their writing habits. From PAD's point of view these events are crucial to better understand the profile of the

writer and how this has affected the way in which his or her working archive was organised.

Particularly challenging is the description of the archives, in the first place for the huge quantity of files and the variety of formats. But the real change, vis-à-vis traditional archives, lies within the fluidity of the digital medium. This accounts for the increased number of intersections among data and more importantly for the way in which files get stratified. Successive deposits of portions of a traditional paper-based archive cause an increase in the number of files, since no files are being replaced. When the same procedure is carried out in a digital environment the results are quite different: some contents could disappear or be replaced, files could be shifted from one folder to another, technical metadata could differ as a consequence of the replacement of the equipment or modifications in the software configuration, and so on.

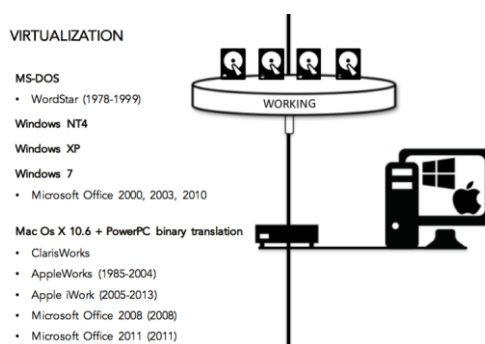
*Fig. 6: Procedures carried out within the Working Area*



Access to data and display is another issue. Two approaches can be followed. The first one is the traditional archival approach, aiming at recreating the original context, that is the organisation that the author him- or herself gave to his or her electronic working papers. The other could be defined as “bibliocentric”, since the individual work – even in an abstract sense – acts as aggregating element for the associated documents: pre-textual material, the various drafts of the text, correspondence and other material exchanged with the publisher, the editor, friends and colleagues, family, more or less complete

editions, work presentations, reviews, readings, readers' comments, intersections with other works, and so on.

**Fig. 7: Emulation. Old softwares run on virtual machines supporting various operating systems.**



Despite the fact that the bibliocentric approach can be seen as more friendly and intuitive for certain segments of readers, it hardly tells us anything about the author's writing habits, the way in which his or her mental, digital, or relational spaces were organised.

Therefore within PAD we are considering the possibility of combining an app for describing the higher levels of the fonds according to the traditional hierarchical structure of an archive with a contents-oriented management system which should enhance the search functionalities. Possibly, the data model should be designed according to the FRBRoo (2015) model, since this appears to be the bibliographic model best suited to organise the individual entities (works, manifestations, linked works, different formats, etc.) within a grid of relationships of a very diverse nature and to give value to events marking the life of the work.

The variety of formats and the great difformity of each personal archive require that the system be designed with a sufficient degree of abstraction, and that the nature and type of these relationships be increased in number as a new instance turns up in the course of cataloguing.

It goes without saying that due to its size there is no way that an archive can be analytically described manually. Specifically designed data mining procedures should provide the cataloguers with raw material which will then undergo a limited number of checking and validation actions, as well as data enrichment whenever necessary. An app aiming at making data entry and validation a straightforward procedure, by offering the operator on the same screen a basic

template and the display of the item (text, photo, audio-video, etc.) he or she is describing, is in development.

Bearing in mind the layering issues and the intersection of contents, PAD should be able to display an archive structured and described along these lines in more unconventional (at least for this area of studies) and yet effective ways than index-based interfaces. Existing relationships among entities should be explicitly and unambiguously shown resorting to tridimensional visualisation, to graphs and to other equally impacting devices.

Eventually it must be emphasized that the thorough exploration of a literary archive cannot be the responsibility of few operators. The system must therefore make available a reasonable number of tools enabling crowdsourcing and the broadest circulation of contents originated from the archives.

Confronted with such huge challenges and aiming at a better knowledge of the literary phenomena interpreted by the authors who have already delivered their working papers, as well as by those who will hopefully deliver theirs in the years to come, PAD would be delighted to offer itself as a partner for reflections on standards and best practices, as a testbed for digital implementations, and as a source of literary contents for researchers and readers.

#### **References**

- CCSDS, (2012). *Reference model for an open archival information system (OAIS): Recommended practice* (CCSDS 650.0-M-2: Magenta Book). Consultative Committee for Space Data Systems Secretariat, Washington, DC. Web: <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- Cragin, M. H. [et Al.], (2007). *An Educational Program on Data Curation*. Graduate School of Library and Information Science University of Illinois at Urbana-Champaign. Web: <http://www.ala.org/acrl/sites/ala.org/acrl/files/content/aboutacrl/directoryofleadership/sections/sts/stswebsite/programs/annual2007programs/annual2007posters/Cragin2.pdf>
- Diamond, D., (2012). *DAM Survival Guide: things to know (and avoid) when designing, promoting and maintaining the perfect digital asset management initiative*. [Seattle?], CreateSpace Edition
- Erway, R., (2010). *Defining “Born Digital”*. An essay. OCLC, Dublin, OH. Web: <http://www.oclc.org/research/activities/hiddencollections/borndigital.pdf>
- Erway, R., (2012). *You’ve Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media*. OCLC, Dublin, OH. Web: <http://www.oclc.org/research/publications/library/2012/2012-06.pdf>
- FRBRoo, (2015). *FRBR object-oriented definition and mapping from FRBR<sub>ER</sub>, FRAD and FRSAD (version 2.2)*. International Working Group on FRBR and CIDOC CRM Harmonisation, Supported by Delos NoE, Editors: Chryssoula Bekiari [et Al.]. Web: [http://www.ifla.org/files/assets/cataloguing/frbr/frbroo\\_v2.2.pdf](http://www.ifla.org/files/assets/cataloguing/frbr/frbroo_v2.2.pdf)
- Kirschenbaum, M. G., Ovenden, R. and Redwine, G., (2010). *Digital forensics and born-digital content in cultural heritage collections*, with research assistance from Rachel Donahue. Council on Library and Information Resources, Washington, DC
- Lavoie, B., (2014). *The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)*. DPC Technology Watch Report 14 (2 Oct. 2014). Digital Preservation Coalition. DOI: <http://dx.doi.org/10.7207/twr14-02>.

Lee, C., and Tibbo H. R., (2007). Digital Curation and Trusted Repositories: Steps Toward Success. *Journal of Digital Information* [Online], 8.2. Web: <https://journals.tdl.org/jodi/index.php/jodi/article/view/229/183>

Lee, C., Tibbo, H. R., and Schaefer, J., (2007). Defining What Digital Curators Do and What They Need to Know: The DigCCurr Project. Proceedings of the *7th ACM/IEEE Joint Conference on Digital Libraries, Vancouver, British Columbia, Canada, June 18-23, 2007*, ed. by R. Larson [et Al.]. ACM Press, New York, NY.

Magan, A., (2005). *Intro to Digital Asset Management: Just what is a DAM?* [blog posted at Real Story Group on 30 April 2005, archived by Internet Archive from the original on 22 July 2012]

TRAC, (2007). *Trustworthy repositories audit and certification: Criteria and checklist (Version 1.0)*. The Center for Research Libraries and Online Computer Library, Chicago, IL and Dublin, OH. Web: [http://www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf)

Van Niekerk, A. J., (2006). *A methodological approach to modern digital asset management: An empirical study*. Allied Academies International Conference (New Orleans, Louisiana April 12-15, 2006). Proceedings of the International Academy for Case Studies, Vol. 13, No. 1

Weston, P. G., (2013). Digital curation: nuove opportunità per bibliotecari e archivisti. *Il libro, gli archivi e la memoria digitale*, a cura di P. G. Weston. Unicopli, Milano, 9-66.

*Workbook on Digital Private Papers* (2008). Web: <http://www.paradigm.ac.uk/workbook/index.html>. Integrated by *Guidelines for creators of personal archives* (2008). Web: <http://www.paradigm.ac.uk/workbook/appendices/guidelines.html>

#### Websites

PAD (Pavia Archivi Digitali) website: <http://pad.unipv.it/>

PADI (Preserving Access to Digital Information) website: <<http://nla.gov.au/nla.arc-10691>>

Paradigm (Personal Archives Accessible in Digital Media) website: <<http://www.paradigm.ac.uk/index.html>>