

Superb Stewardship of Digital Assets – Developing a strategy for Digital Archiving and Preservation at the University of Notre Dame

H. Hockx-Yu
Office of Information Technologies,
University of Notre Dame
Notre Dame, IN 46556
USA
hyu3@nd.edu

ABSTRACT

This paper describes the effort to develop an institutional strategy for digital assets management for the University of Notre Dame, where information is currently managed within a devolved organisational structure, and archiving and preservation are often overlooked. It advocates a lifecycle approach of digital assets management and recommends a strategy with the goal to embed considerations for archiving and preservation in policies, workflows and technologies across the entire organisation.

CCS CONCEPTS

• **Applied computing** → Education → Digital libraries and archives

KEYWORDS

Digital Archiving, Digital Preservation, Digital Assets Management, Strategy Development, Institutional Challenges

1 INTRODUCTION

Just a few years ago, many of the materials universities used for teaching, learning, research and communications existed in physical formats. More and more of this analogue world now presents only as digital artefacts. Much of the learning, teaching and research at today's universities is done digitally. Digital records and data are now ubiquitous and their proper management and preservation is critical for academic institutions to achieve their overall aspirations.

The University of Notre Dame (UND) is a private research university located adjacent to South Bend, Indiana, United States. It was founded 1842 by Father Edward Sorin, a priest of the Congregation of Holy Cross. UND is ranked #15 among National Universities in 2017 [19] and #12 of the top research universities in 2016 [2]. Catholicism and Catholic intellectual tradition continue to serve as the underlying foundation for the University's aspirations as a community of scholars and administrators. It is also the University's point of greatest distinction from many other research institutions [15].

UND's leadership consists of Fellows, Board of trustees, Deans of colleges and schools and the President's Leadership Council. As with many higher education institutions, while there is a central process of allocating funds, devolved managerial responsibilities and budgeting are essential elements of the management model. Authority and accountability are with business units, including the autonomy of spending budgets, regardless of the sources: whether it is centrally allocated funds, return on endowment or grants. The organisation and structure of e.g. IT services in this context consist of a hybrid of practices. While the Office of Information Technologies (OIT) has the responsibility of supporting enterprise-wide computing on campus, it merely serves as a technology partner or a service provider. Business units are fully autonomous in deciding whether to use OIT provided services or solutions. Some business units for this reason have their own IT team to ensure support that is more sensitive to local knowledge or requirements. IT infrastructure services, however, is an area where high rate of adoption can be observed and is increasing. There is in general a good collaborative relationship between the OIT and local IT departments.

As digital information profoundly changes learning, teaching and research, it also brings about many challenges, ranging from discovery among a vast amount of information, diversity of formats, and rights management to analytics. The biggest challenge is however organisational, how the University as an organisation in its entirety ensure consistency or achieve economy of scale, being aware that devolution tends to encourage duplication or creating silos.

Since 2009 various initiatives have already taken place at UND to coordinate efforts and address the challenges of digital assets management. They were of different scale and involved different academic and administrative units. A taskforce or project based model has been commonly used to support these efforts, where executive-level committees were formed to oversee the initiatives, who tasked working groups with participants across campus units to document and prioritise requirements and make recommendations. These initiatives have put in place pieces of infrastructure and some processes needed for digital assets

management and placed NUD in a good position to revisit the topic and address the gaps.

Between October 2016 and March 2017, the author undertook a consultation exercise involving a wide range of stakeholders across campus, including interviews and meetings with over 80 individuals from over 20 departments. The ultimate goal was to develop a campus-wide strategy to guide the preservation, access and dissemination of NUD’s digital assets. The findings and recommendations are reported in this paper.

2 DEFINITIONS

In the context of NUD, the term “digital assets management” is often used to refer to the activities related to managing digital information. The use of the word “assets” implies an underlying process of making value judgement. Regardless of the formats, digital assets are records, data and resources, typically owned by the University, regarded as having value and therefore needing to be retained and/or preserved. The implied valuation raises questions with regard to the identification and appraisal of digital information, as well as roles and responsibilities: how does the process of attaching value work and whose responsibility is it?

Three broad, not necessarily mutually exclusive categories have been identified to scope the digital assets landscape at UND:

University Records

The Records Management and Archives Policy [18] defines University Records as “recorded information created or received in the course of conducting University business and kept as evidence of such activity. This definition extends to records in all formats, including but not limited to e-mail, electronic and paper documents, film and print graphics, and audio and video recordings.

Research Data

The Data Retention and Access Policy [9] defines Research Data as “information recorded or customarily recorded in the relevant field, as a result of research. Data include notes, records, slides, photographs, drawings, information stored in electronic and/or computer readable form, reports, publications, correspondence, and summaries, compilations, or derivatives of other data.

Resources for teaching, learning and research

This category of digital assets refers to the wide range of content the University collects, creates, receives as donation, or purchases to support teaching, teaching and research. Examples include the Library collections, donated datasets, electronic databases, electronic journals, also lecture captures and Massive Open Online Courses (MOOCs).

The lifecycle model [1] is a commonly accepted framework for managing digital information and data, which recognises the change of value and frequency of use over time. It also allows the identification of different stages in the lifespan of digital assets

from creation to archiving that otherwise might be overlooked. There are different ways of defining the stages of digital assets, dependent on institutional practices and the level of detail. The model outlined in Figure 1 is a generic and broad one, omitting much detail yet applicable to all categories of digital assets created and acquired by UND. The key is to apply this way of thinking systematically and to make sure that strategies and policies are consistent, and the crucial stage of “Archiving and Preservation” is not routinely overlooked.



Figure 1: Digital Assets Lifecycle

For the purpose of the work described in is paper and based on UND’s specific circumstances, the terms “digital archiving” and “digital preservation” are defined as follows:

Digital archiving

The process of acquiring and appraising records and data for long term retention. It also includes all the activities related to providing access to archived data and records.

Digital Preservation

The series of managed activities that ensures ongoing access to digital material, with a focus on guarding against deterioration and technological obsolescence over time.

2 METHODOLOGY

The consultation was conducted by the author, in her role as Program Manager for Digital Product Access and Dissemination, reporting to the UND Vice President for Information Technologies & Chief Information and Digital Officer. The author’s post is jointly supported by the Office of Information Technologies and the Hesburgh Libraries.

The goal of the consultation was to gain understanding of the current state of digital assets management across campus, and to identify the gaps. The findings will inform the development of a campus-wide strategy and a program of work for its implementation.

Over 80 individuals were consulted across campus, mostly through face-to-face meetings and interviews. The interviews were semi-

structured, using predefined as well as open questions. The focus was on obtaining an overview of the current practice and understanding what happens to digital assets throughout the lifecycle: what digital content are created, how they are managed, used, archived and preserved.

A small group of individual researchers, who do not necessarily represent a college or department, were also consulted. Their requirements include end-of-project data management, preservation of personal research material and support for collaborative research. One of these is an anthropologist, whose personal research “archive” spanning over 25 years contains analogue and born-digital material of various formats including human subject data.

When use cases were identified for collaboration, or referral of (otherwise unknown) services, efforts have also been made to bring people together across business units and functional teams. This provided an opportunity to think about future workflows and processes based on real-world problems and already led to better collaboration, potential grants and improved processes and access to some digital assets. Selected use cases are described in Section 4.

The main goal of consultation exercise was to obtain a broad picture of digital asset management at UND and therefore had an internal focus. Benchmarking was an element of the project and served as point of reference, rather than motivation for change. It was mainly based on comparison with similar research institution in the US and in the Midwest, including Stanford University, Yale University, Princeton University, University of Michigan and Indianan University. The National Digital Stewardship Alliance’s digital preservation levels were also used to assess UND’s compliance [8]. This pointed to e.g. the lack of dedicated positions related to digital preservation, and some NDSA level one data protection issues.

Research reported in this paper is based on data collected from a large number of individuals. Caution has been applied to make sure research ethics have been taken into account and the appropriate protocols are being followed. A formal compliance review submission was made to UND’s Research Compliance, and the Institutional Review Board (IRB) determined that the project qualifies as exempt human subjects research, as its nature does not meet the regulatory definition.

3 FINDINGS AND RECOMMENDATIONS

In this section, findings and recommendations that emerged from the consultation exercise are reported. The findings describe the current digital assets management landscape at UND. The recommendations directly respond to the findings and suggest specific interventions or strategies to address the identified gaps and challenges.

3.1 Findings

While previous digital assets management initiatives have delivered solutions or capability, these tended to meet specific requirements and did not necessarily achieve maximum adoption at the institutional level. Many recommendations were not followed through or resources for implementation were not in place. The executive committees and taskforces were charged with understanding requirements and making recommendations, but not with the overall implementation or change management, leading to start-stops over time.

Finding 1:

Previous digital assets management initiatives did not make enough impact at the institutional level, leaving many valid recommendations unimplemented.

While there are excellent examples of active collaboration between groups on campus, a common feedback on the current practices of digital assets management points towards a “siloeed” approach, referring to uncoordinated and different ways of doing things in overlapping areas. Individual business units continue to create and manage their own digital assets. This has obvious disadvantages:

- Digital assets are stored at different locations, lacking an overview at the University level
- Difficulty to discover knowledge of and to locate digital content, resulting in poor reusability
- Duplication of effort, leading to the proliferation of a large number of applications, which are costly to support and difficult for users to choose from
- Negative effects on collaboration
- Difficulty to develop common standards and encourage best practice.

Administrative hurdles were also cited. Some stakeholders are of the view that the tendency toward intramural charges and cost recovery fees to internal users erects economic barriers and (inadvertently) encourages silos. This has also been attributed to the uneven level of use of technology across the organisation. While some departments engage with the most advanced technology, others still operate in IT environments typical of the 1990s.

Finding 2:

Charging models for services are inconsistent across campus, and can form barriers for consistent lifecycle management of digital assets.

Finding 3:

Intended enterprise solutions are often unknown, not considered or not affordable to SOME departments.

Lifecycle management of digital assets, especially archiving and preservation arrangements, needs to be embedded across the organisation, reflected in institutional policies, information and data governance, IT services management and day to day business processes and workflows. It needs to be taken into account from

the point of creation and considered especially when entering into contracts with 3rd party suppliers. This is currently not the case. Archiving and preservation are routinely overlooked.

Finding 4:

There is generally a strong focus on “now”. Risk of digital assets being lost is not well understood, nor what is required to ensure their continued access.

The amount of digital content we create far exceeds the amount we are able to keep and preserve. Archiving is as much about keeping as it is about discarding things. Common questions being asked in this context are what do we keep and who should decide on this? There are a number of university-level policies which define retention periods for records and research data, and go a long way in help answer these questions:

Table 1: University Policies relevant to Records and Data Retention

Policy Name	Responsible Department	Designated Custodian	Review Date
Data Retention and Access [9]	Research Administration	Faculty member in charge of the research or Non-faculty principal investigator or other person designated by the relevant department chair or centre director or Department chair or centre director	
Intellectual Property Policy [13]	Office of the Provost		October 2015
Retention of Grant or Contract Financial Records[20]	Research & Sponsored Programs Accounting	For electronic transactions: the central administrative office responsible for maintaining the electronic application For paper documents: the office receiving the original document	August 2012

The Records Management and Archives Policy [18]	University Archives	University Archives (in collaboration with the Office of General Counsel and University administrative offices)	2015
General Records Retention Schedules [16]			

Finding 5:

There is a lack of general awareness of the University’s data and records retention policies. And it is not evident how these are implemented for assets in digital formats.

Some departments are not fully aware of the role and expertise of the Hesburgh Libraries and University Archives. Curatorial decisions are made without consulting them; records are created or collected and kept locally (despite having become inactive) without depositing to the Archives for extended period of time. Some of these records are stored on obsolete media or in obsolete formats, which are at risk of being lost without immediate preservation actions.

Finding 6:

University Archives’ responsibility for collecting, maintaining and preserving the official records of the University is not well understood. It is not commonly accepted that this responsibility extends to include digital records.

Although digital archiving and preservation are not alien concepts, there is a lack of technical skills and practical experience at the institutional level. Without detailed retention schedules and identification of systems of records, archiving requirements are rarely implemented in the many systems and applications that are used to handle digital documents, records and data. Basic bit-level preservation actions including periodical fixity checks are not explicitly or systematically carried out. Digital files are often deposited or transferred using physical media, without essential metadata or checksum-based verification. The Hesburgh Libraries and University Archives, who have the long-standing mission of preserving knowledge and institutional memory, are still in the process of analogue to digital transition. They are working towards active preservation of collection items and systematic collection of institutional records in digital formats, both requiring time, investments and development of relevant skills.

Finding 7:

Some digital assets have already been lost. Some are currently stored on obsolete media or in obsolete formats, not accessible and at risk of being lost, without (immediate) remedy actions.

Finding 8:

Born-digital University records are a key area of challenge for the University Archives. An up-to-date, function-based retention schedule (as opposed to office-based), and identification of system(s) of records will allow wider application and support automation. However, without the skills, resources and tools required to acquire, appraise, preserve and provide access to these records, it is very challenging for University Archives to fulfil their role.

Research data takes many forms. Archiving and preservation of research data, because it covers a wide range of disciplines with diverse data management, sharing and curation practices, is a complex area where there is no one-size-fits-all solution. Many universities approach the challenges by setting policies, putting in place storage and data repositories and by providing support and guidance for data management - an important element is to meet public access requirements for federally funded research. The ultimate goal is to support the key principle underpinning the scientific discourse: the ability to verify results and claims independently, which is being seriously jeopardised by the disappearance of referenced data in research articles as well as the reliance on computer hardware, software and configurations and source code.

Notre Dame Research, various computing and data centres, as well as the Hesburgh Libraries already offer a wide range of services and support to researchers. A good example is how grants and awards are managed, where the University's system of record for research administration, Cayuse, plays an important role. We however need to question if individual researchers can be expected to preserve research data reliably in the long term. If yes, what else needs to be put in place to support them fulfil this role.

CurateND is UND's digital repository committed to providing permanent access to and long term preservation of the University's Research and Scholarship, including published work and research data. It also serves as the Hesburgh Libraries' digital library system. CurateND is a relatively new service, only launched in 2015, and the team has so far focused on the technical development, based on identified use cases. CurateND is not solely an open access repository. It contains open access, embargoed, and private content. The positioning of CurateND is to become the system of records for the University's scholarly outputs, regardless where the data resides.

CurateND has a set of policies defining various aspects of the repository, including a Review and Retention Policy [4]. The policy commits to retaining certain types of material in perpetuity and defines the retention periods and review criteria for the remainder of the content.

It is still early days, the deposit experience with researchers however seems somewhat mixed. While many researchers understand the potential risks of information and data loss, a

degree of reluctance has also been observed. This could be due to a number of reasons:

- Different disciplinary practices in data sharing
- Lack of trust in ceding control of data
- Unclear about ownership and rights
- CurateND does not meet specific requirements
- Depositing research data seen as "one more thing to do"
- Lack of awareness of the existence or purpose of CurateND.

Finding 9:

The association between research data and administrative records are not maintained. The dual role of CurateND as the institutional repository and Hesburgh Libraries' digital library system is intended to optimise use of the technology. However, it does at times create competition in terms of resource as well as priority.

Storage and backup services are crucial to protecting digital assets. A number of enterprise storage options are offered by the OIT, serving different purposes. Some are unlimited and funded centrally; some come with a base entitlement and charge a fee for additional space. Many departments also use locally sourced storage solutions, e.g. RAID, SAN, cloud storage, due to access requirements or budget restrictions. Key observations and feedback specific to storage include:

- Data is often inconveniently spread across different storage locations, without the confidence or confirmation of proper back-up.
- Different storage solutions are not well understood; some departments experience shortage of storage capacity.
- Alternative storage solutions, e.g. network-attached storage (NAS) devices and external disks, are commonly used, not just as short relief. These do not scale cost-effectively, are not intended for long-term use and can lead to data loss.
- The longevity of some storage solutions (e.g. Google Drive) are unclear.
- Individual researchers' desktops are not consistently backed-up.
- Some block-level storage solution, e.g. the LTO Tape Archival Storage, is not widely known on campus.

Finding 10:

Despite the existence of enterprise storage solutions, much data still seems to be stored on (scattered) direct-attached storage, or locally managed NAS devices, creating data islands and poor data protection.

3.2 Recommendations

The recommendations are summarised in this section, grouped into a number of areas.

The longevity of our digital assets is a campus-wide challenge. While business units can take actions on their own there is a strong

consensus that coordinated strategies and shared resources are preferred over siloed local efforts.

Two particular frameworks are recommended to guide and benchmark UND's digital archiving and preservation efforts:

- "Levels of Digital Preservation", developed by the National Digital Stewardship Alliance [8].
- "Research Data Storage: A Framework for Success", developed by the EDUCAUSE Center for Analysis and Research [5].

3.2.1 Strategy, Policy and Organisation

UND has an excellent track record in managing paper records and physical assets such as buildings. As more and more assets exist in digital formats, we must apply the same rigor demonstrated in planning and management and take the same coordinated approach to managing our digital assets, starting with a recognition of their value and their stewardship as a strategic priority. One of the University's strategic goals is to "foster the University's mission through superb stewardship of its human, physical, and financial resources" [12]. It is pertinent to also include the vast amount of digital resources we create, procure and are given, that underpin our academic and administrative activities, and aim to provide a track record of equally excellent stewardship for this new class of University assets.

Recommendation 1:

Amend the 4th goal in the University's Strategic Plan to include "digital resources" as an additional area where superb stewardship is required, so that it reads: foster the University's mission through superb stewardship of its human, physical, financial and digital resources.

Recommendation 2:

Develop a business model so that the required resources are in place to prioritise and implement the recommendations made in this report.

Stewardship of digital assets is not a one-off event but an ongoing process, operating in tandem with the full range of services supporting digital information environment and the overarching financial and organisational infrastructure. Project or taskforce-based efforts are time-bound, focus on specific outcomes, so may not be appropriate for core activities such as digital stewardship, which requires processes proceeding continually over time. This does not mean that we should not ever run projects and task-forces to solve specific problems. The key is to emphasise programmatic progress so that outcomes delivered by projects and taskforces are synthesised, followed up and embedded in business processes through change management.

Recommendation 3:

Move from project and taskforce-driven activities to a fundamental programme of core activities by establishing a permanent organisational fixture (e.g. a person, an office or a team) with

campus-wide remit to coordinate policies, procedures, practices, resources and technologies related to digital assets management.

Digital resources are far from being unmanaged at UND. A diverse array of people, technology and processes are in place, forming the environment in which digital resources are created, stored, secured and used. What we do not have is a culture or the practice of thinking beyond now, i.e. what happens to digital resources beyond the stage of active use? What must we do now to ensure we can find and interpret our digital records and research data in the future? These questions cannot be ignored until digital resources are passed into the custody of the Libraries and Archives, nor does the responsibility of digital preservation reside solely with them, only at the very end of digital assets' lifecycle.

Recommendation 4:

Carry out campus-wide effort to identify the process and control points where digital archiving and preservation considerations can be routinely embedded, e.g.

- *Campus-level information committees and governance bodies*
- *Service units creating digital resources and departments commissioning the work*
- *Procurement including 3rd party data / information creation and hosting*
- *IT service management framework*
- *Business and project planning*

3.2.2 Library and Archives

Among the three categories of digital assets defined in section 2, resources for teaching, learning and research is a category for which the University Library and Archives have shared responsibility of long-term custodianship. Some archival records have research values and can be used as education resources, and vice versa. As technologies advance and new types of resources are created and added to the category, it is important that the Library and Archives keep abreast of new developments and clearly define roles and responsibilities, to avoid grey area and to ensure none of the assets fall between the cracks.

Discovery has been made of digital assets stored on obsolete media or in obsolete format (see examples in 4.1 and 4.2). Some of the legacy assets are collection items held by the Libraries and Archives, some belong to researchers' personal archives of research data. They are no longer easily accessible and are at risk of being lost without immediate remedy action. There are also analogue materials stored on physical media which need to be digitised to avoid media deterioration.

Digitisation and format migration routinely take place at the Libraries and Archives. In addition, a number of academic and teaching support units also have old equipment and offer faculty and students a similar service (which are not extensively used and the usage is declining). It makes sense to consolidate these as a campus-level service.

Legacy content is not the sole challenge. As an institution we also need to build the capability and expertise to preserve current and advanced digital resources. Novel classes of born-digital data, information and knowledge are being created, each with its own underlying artefacts needing to be preserved. Complex digital objects such as augmented reality, mobile apps and gigapixel images are increasingly used to support teaching, learning and research. We collectively do not yet know the best methods for archiving and preserving them. We need to take actions to address the challenges, together with peer institutions and the digital preservation community. We need to invest in a permanent team within the organisation who can build the know-how and technical expertise for hands-on and ongoing digital preservation activities.

Recommendation 5:

Establish a digitisation and digital preservation centre, based at Hesburgh Libraries, to serve campus-wide needs. The centre will digitise, preserve, research, and develop solutions for preserving advanced digital resources. It will also educate and advise, by recommending and helping to implement common standards and encouraging best practice.

Some institutional records reside in systems and locations external to the University, e.g. on the web and social network platforms, without our active control. This type of records is also ephemeral in nature and can easily disappear - the commonly cited average lifespan of web pages is about one hundred days. Collecting or archiving these records has therefore become a routine activity for many organisations including over 250+ academic institutions in the US, including all the Ivy League schools. Currently UND's web and social presence are not systematically archived.

Recommendation 6:

Hesburgh Libraries and University Archives need to make an informed decision on whether to archive the University's web and social network presence as institutional records. Also consider the potential of web archiving in helping to build digital collections in areas where we have strong existing physical collections or research interest, e.g. Catholicism, Irish studies.

Born-digital records are the most challenging issue for University Archives. There is a recognition that the records management function in the Archives needs to prioritise and focus accelerated attention on electronic records to ensure the University's history, its legal and administrative records are documented appropriately. Efforts are already underway to develop function-based retention schedules and to put in place dedicated electronic records management resources.

Recommendation 7:

Build the capability for and more rapidly transition to electronic records management. Archives should work closely with the technological partners on campus to identify systems of records and find ways to implement the appropriate retention schedule in

these systems. As a minimum, disaster recovery plans must be in place for vital records.

3.2.3 Research Data

Archiving and preserving research data is a complex challenge which require collaboration between many campus units and with individual researchers. To progress in this area, the best strategy is to build on the work already done internally by Notre Dame Research, various computing and data centres, Hesburgh Libraries as well as academic departments, and externally by various research and disciplinary communities. A two-pronged approach is recommended: at the institutional level, meeting public access requirements of sponsored research is no-doubt a priority. From individual researcher's perspective, an immediate task would be to discover and "rescue" research outputs which have become inaccessible due to technology obsolescence – this should one of the services offered by the digitisation and digital preservation centre mentioned in Recommendation 5.

Recommendation 8:

Review, update and coordinate the research data retention policies including explicit treatment beyond the stipulated retention periods. Make sure the designated custodians are aware of and are supported in fulfilling their roles. If not already in place, develop the necessary procedures to implement these policies.

Recommendation 9:

Consider measures that can help ring-fence the two distinct purposes of CurateND, as UND's institutional repository and Hesburgh Libraries' digital library system, to ensure that requirements for both are met. These could be financial, organisational or branding-related (e.g., dedicate more resources to each pipeline of content).

Recommendation 10:

Maintain central records of research outputs (publication and data) arising from (at the very least) sponsored research, regardless of whether these are uploaded to CurateND, disciplinary repositories or other repositories required by funders, and link these with the corresponding administrative records held in the Cayuse system.

3.2.4 Storage and Cloud Services

Sitting low in the technology stack, storage and backup are the foundation and most important enablers for any archiving and preservation endeavour. Arguably a centralised approach has a lot of advantages. This, for example, will allow identification / inventories of digital assets and enable bit-level preservation actions. The findings around storage however points to the need for basic steps to ensure data protection, e.g. by reducing the use of direct-attached devices as a long term storage solution, and ensuring data is backed-up appropriately. We also need to communicate effectively about the purpose and cost structure of

enterprise storage solutions. If departments cannot use enterprise solutions, either because of performance or funding issues, an alternative approach, instead of driving people to implementing local solutions, is to work with departments and find ways to offer a service that DOES meet their requirements.

Just as many other higher education institutions, Notre Dame has widely adopted the Software-as-a-Service (SaaS) approaches to Cloud computing and are transitioning significant portions of our infrastructure to the Cloud [10]. This Cloud First effort is guided by a lifecycle framework, which includes an “end of life” stage of a solution or a service [11]. Explicit treatment of data, when retiring a service or a solution, however will need to be added and defined, based on the appropriate retention schedule, to make sure data ends up in the right places at the end of the lifecycle.

80% of UND’ IT services are expected to move into the Cloud by the end of 2017. We are aware of the risks associated with cloud computing such as information security and regulatory compliance. Moving data to the cloud however also has implications on its longevity and there are a number of specific aspects we need to consider:

Data integrity

How do we manage the integrity of stored content over the duration of the service, which encompasses not only protecting data from unauthorised alteration, but also from bit rot. How can we ensure data is safeguarded in the event of disasters, supplier failure, or a decision to change service provider? In addition, what happens to the data when a service is retired, or if a supplier goes out of business, and in the event of a contractual dispute or termination?

Archiving and preservation

It will take some thought to work out how to archive and preserve (inactive) data sitting in the Cloud, and how these processes integrate with the many applications holding and processing the data in its current lifecycle. The challenges are in fact parallel to doing this on campus. The location of data does not change the nature of archiving and preservation activities. Many vendors already offer cloud services for archiving and preservation, with the ability to integrate with key cloud storage platforms.

Recommendation 11:

Review storage service strategy in light of the findings. Make plans and take actions to address the identified issues. An immediate goal should be reducing the use of direct-attached devices as a long term storage solution, and ensuring data is backed-up appropriately.

Recommendation 12:

Identify long-term preservation risks related to Cloud services and put in place measures to mitigate the risks (e.g. through contractual agreements). Explicitly define treatment of data, based

on the appropriate retention schedule, such as when retiring a service in the Cloud.

4 USE CASES

The section offers description of a number of selected use cases, which came to the author’s attention during the consultation exercise. These real world examples support the findings and recommendations described in Section 3. They also illustrate the ongoing nature and urgency of digital archiving and preservation.

4.1 Recordings at risk

An internal report delivered in 2010 by one of the previous digital assets management taskforces mentioned an archive of historically significant sound recordings held by WSND, the FM radio station that is part of UND’s Student Media Group. These recordings are on open reel tapes and stored in an office environment. They include interviews with prominent campus figures such as Father Hesburgh and speeches given on campus by Robert Kennedy and a number of presidents. A recent inventory-check also revealed recordings belonging to the Notre Dame Sophomore Literary Festival (SLF), a student-organised event that invited the top names in literature to give readings, deliver lectures, and engage in panel discussions on campus, particularly in the event’s early years in the 1960s and ’70s. Twenty-five poets, fiction writers, dramatists, and critics came to Notre Dame to participate in the SLF. Prominent among the participants were Allen Ginsberg, Robert Duncan, Jerzy Kosiński, Tom Stoppard, Ishmael Reed, John Barth, Diane Wakoski, and William Gass.

Approximately 45 hours of SLF readings and lectures on sixty-eight open reel ¼” tapes from 1968, 1971, 1972, and 1979, currently housed at two separate locations, partially by WSND and partially by University Archives, were selected and brought together, which formed the basis of UND’s grant application for the Council on Library and Information Resources’ (CLIR) Recordings at Risk Program [3]. As a result of the project, the tapes housed at WSND will be transferred to the University Archives for optimised storage and better access, and added to the Archives’ online finding aid.

4.2 1966-1967 Sisters Survey

Sister Marie Augusta Neal was professor of sociology at Emmanuel College in Boston, Massachusetts between 1953-1991. She became the Director of the Research Committee of the Conference of Major Superiors of Women’s Institute (CMSW) which conducted the CMSW Sisters’ Survey of 1966-1967, a population attitude survey designed to assess American sisters’ readiness for renewal. With 649 variables and responses from over 130,000 Catholic sisters, it is believed to be “the largest, single, data gathering event ever performed with regard to women religious” [7]. All materials related to the CMSW Sisters’ Survey and Sr Neal’s follow-up studies were donated to UND Archives between 1991 and 1996, including computer data tapes [17].

The Sisters Survey data was initially assembled, processed and stored on IBM EBCDIC format tapes and converted to newer formats in 1996, before being deposited to the UND Archives. Under the custodianship of UND, the survey data was transferred to CDs then to computer hard disk in 1999. The Survey data has fortunately survived the format migrations but has not been used for 18 years since 1996.

The data came to the author recently without any file extension. Running the file through DROID [21], a file profiling tool developed by the UK national Archive did not deliver any useful information. Knowing that we are looking at cross-tabular data, and the possible involvement of statistical software, a few obvious extensions were added. The files could be opened with Excel but contained endless lines of numbers without much meaning so the first challenge is knowing how to truncate the data.

The truncation problem was later resolved by UND’s Economics Librarian, who has a good understanding of survey data – “I just saw the pattern”, he explained.

Things followed smoothly from this point onwards. The dataset has now been reformatted and stored in .dta and .csv formats. The “codebook” was also recreated – our librarian wrote scripts and extracted all the questions and pre-defined responses from PDFs and pulled them together in one document. The dataset is in the best possible format for re-use. We are just dotting the i’s and crossing the t’s, before releasing it publicly.

A key takeaway of the above use case is that just keeping the bits safe is not simply not enough. Active use is the best way for monitoring and detecting digital obsolescence. Materials created by obsolete technology require ongoing care and necessary intervention such as media and format migration. Furthermore, without metadata, in this case the notes, finding aid and scanned codebook, it would be very difficult to interpret or make sense of the Sisters Survey dataset.

4.3 Deposit workflow for University Archives

It came to the author’s attention that the current workflow for depositing or transferring digital files to University Archives relies heavily on the use of physical or portable media such as DVDs, external hard disks and flash drives. An example of this is how UND’s Debartlo Performing Arts Center (DPAC) deposit recordings of concerts and performances. Recordings are initially stored on DPAC’s network drive. Once a year these are burnt to DVDs, to make space for new recordings and for deposit to the University Archives. The workflow was set up at a time when network bandwidth was inadequate for transferring large audio files.

A small project took place to improve the workflow and take advantage of current network capacity. The Storage Services team in the OIT has now set up a dedicated S3 bucket to allow individual departments to drop files for the Archives. Those

regularly deposit files to the Archives no longer need to copy files to physical media, and instead place the files in the department’s folder on S3. The Archives will move and process the files, and store them eventually in the Spectra T950 Tape Library, UND’s enterprise archival storage service. Archives can also use the same folder to deliver back any archival files requested by the departments for re-use.

Additional workflows being developed and hinge on this new workflow include file naming conventions, formats and metadata requirements, and checksum base file transfer verification.

5 CONCLUSIONS AND NEXT STEPS

Digital Archiving and preservation are important aspects of digital assets management and are yet to be systematically approached at UND. Work reported in this papers constitute steps towards achieving superb stewardship of the University’s assets which now include digital assets.

UND is currently undertaking the Campus Crossroads Project, the largest building initiative in the history of the University [14]. The \$400 million project is aimed at maximising the potential of one of the most recognizable and centrally located buildings on campus: Notre Dame Stadium. The project includes a digital media centre with a 2,000-square-foot studio, bringing together currently dispersed media production work across campus. This is an opportunity to develop a common solution for managing video assets, including digital archiving and preservation requirements – a separate project is already underway, involving departments across campus who produce, use and provide custodianship over videos.

Stakeholders who participated in the consultation exercise were invited back to a follow-up workshop in April 2017, where findings and recommendations described in the paper were reported. Facilitated discussions also took place with the goal to arrive at a prioritised list of recommendations, and to identify the key barriers associated with their implementation. The results, based on the MosCoW and Kano [6] data collected via a pre-workshop survey and group discussions during the workshop, pointed to recommendations 1, 2, 5, and 7 as highest priorities, followed by recommendation 11 and 12.

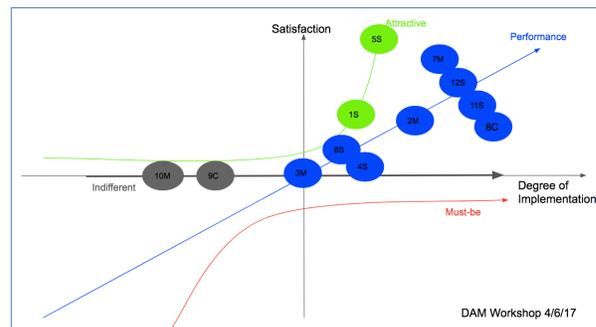


Figure 2: Prioritisation of Recommendations

A further process of prioritisation and business planning process will follow, to design a programme of work to implement the recommendations, which will be constructed upon the three pillars of policy, process and technology. Digital archiving and preservation are central elements of our forward strategy, as they can help us effectively retain, manage and leverage our digital assets.

REFERENCES

- [1] Bill LeFurgy. 2012. Life Cycle Models for Digital Stewardship. *The Signal*. The Library of Congress. <https://blogs.loc.gov/thesignal/2012/02/life-cycle-models-for-digital-stewardship/>
- [2] College Raptor. 2016. The 25 Best Research Universities – 2016 University Rankings. <https://www.collegeraptor.com/college-rankings/best-research-universities>
- [3] Council on Library and Information Resources (CLIR). 2017. Recordings at Risk. <https://www.clir.org/recordings-at-risk>
- [4] CurateND. 2016. Retention and Review of CurateND Policy. <https://curate.nd.edu/policies/retention-review>
- [5] EDUCAUSE Center for Analysis and Research. 2014. Research Data Storage: A Framework for Success. <https://library.educause.edu/~media/files/library/2014/7/cwg1403-pdf.pdf>
- [6] John McIntyre. 2016. Moscow or Kano, how do you prioritize. <https://www.hotpmo.com/blog/moscow-kano-prioritize>
- [7] Marshal H. Mercer. 1995. Correspondence in Marie Augusta Neal Papers (NEA). University of Notre Dame Archives (UNDA), Notre Dame, IN 46556.
- [8] Megan Phillips, Jefferson Bailey, Andrea Goethals, Trevor Owens and Bruce P. Douglass. 2013. In *Proceedings of the Archiving (IS&T) Conference*, April 2013, Washington, DC. http://ndsa.org/documents/NDSA_Levels_Archiving_2013.pdf
- [9] Notre Dame Research Administration. Date Unknown. Data Retention and Access Policy. https://research.nd.edu/assets/188265/data_retention_and_access_f01.pdf
- [10] Office of Information Technologies, University of Notre Dame. 2017. Cloud First. <https://oit.nd.edu/cloud-first/>
- [11] Office of Information Technologies, University of Notre Dame. 2017. Notre Dame's Cloud Solution Lifecycle. <https://oit.nd.edu/cloud-first/cloud-lifecycle/>
- [12] Office of the President, University of Notre Dame. 2017. A Legacy Expanded, a Strategic Plan for Notre Dame. <http://strategicplan.nd.edu/university-plan/>
- [13] Office of the Provost, University of Notre Dame. 2015. Intellectual Property Policy. <http://policy.nd.edu/assets/203061/intellectualpropertypolicy.pdf>
- [14] University of Notre Dame. 2017. Campus Crossroads Project. <http://crossroads.nd.edu/>
- [15] University of Notre Dame. 2016. This is Notre Dame 2015-1016. <https://www.nd.edu/assets/docs/this-is-notre-dame-2015-16.pdf>
- [16] University of Notre Dame Archives. 2014. General Records Retention Schedules. <http://archives.nd.edu/records/general.htm>
- [17] University of Notre Dame Archives. Date unknown. Marie Augusta Neal Papers (NEA). <http://archives.nd.edu/findaids/ead/xml/nea.xml>
- [18] University of Notre Dame Archives. 2015. The Records Management and Archives Policy. https://policy.nd.edu/assets/185265/records_management_archives_2015.pdf
- [19] U.S. News. 2017. National Universities Rankings. <https://www.usnews.com/best-colleges/rankings/national-universities>
- [20] Research & Sponsored Programs Accounting, University of Notre Dame. 2012. Retention of Grant or Contract Financial Records. https://policy.nd.edu/assets/185270/retention_of_grantor_contract_financial_records_2012.pdf
- [21] The UK National Archives. File profiling tool (DROID). <http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/>