**Elevator Pitch Exercise**

**To whom you are pitching?**

**What is the value of the collection?**

**What problems do you face with the collection?**

**What work needs to happen?**

**How long it will take?**

**What benefits will flow from the work?**

**How does it relate to strategic objectives?**

**What will success look like?**

**Digital Preservation Policies Exercise – Policy Elements**

|  |  |
| --- | --- |
| **APARSEN** | **JISC** |
| 1. Reason and Target Group
2. Type of Data Addressed
3. Sustainability of Funding
4. Roles and Responsibilities
5. Cooperation with Others
6. Availability of Policy
7. Technical Requirements
8. Standards to be Used
9. Access to Data
10. Copyright and Licencing
11. Place in Organisational Activities
12. Risk Assessment
13. Incentives to Participate
14. Policy Updates
15. Preservation Timeframe
 | 1. Principle Statement
2. Contextual Links
3. Preservation Objectives
4. Identification of Content
5. Procedural Accountability
6. Guidance and Implementation
7. Glossary
8. Version Control
9. Financial and Staff Responsibility
10. Intellectual Property
11. Distributed Services
12. Standards Compliance
13. Review and Certification
14. Auditing and Risk Management
15. Stakeholders
16. Preservation Strategies
 |
| **SCAPE** | **Library of Congress** |
| 1. Authenticity
2. Bit Preservation
3. Functional Preservation
4. Digital Objects
5. Metadata
6. Rights
7. Standards
8. Access
9. Organisation
10. Audit and Certification
 | 1. Access and Use
2. Accessioning and Ingest
3. Audit
4. Bibliography
5. Collaboration
6. Content Scope
7. Glossary/Terminology
8. Mandates
9. Metadata or Documentation
10. Policy/Strategy Review
11. Preservation Model/ Strategy
12. Preservation Planning
13. Rights and Restriction Management
14. Roles and Responsibilities
15. Security Management
16. Selection/Appraisal
17. Staff Training/Education
18. Storage, Duplication, and Backup
19. Sustainability Planning
 |

|  |  |
| --- | --- |
| **Policy Heading** | **Statements** |
|  |  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |
| **Policy Heading** | **Statements** |
|  |  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |

|  |  |
| --- | --- |
| **Policy Heading** | **Statements** |
|  |  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |

|  |  |
| --- | --- |
| **Policy Heading** | **Statements** |
|  |  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |

**Scenario: The War Historian's Personal Archive**

**The Scenario:**

A famous academic and war historian has donated his entire personal collection to your archive, including his desktop computer. The majority of his records are on this computer’s hard disk.

**The Challenge:**

You must decide how to assess and preserve the information on this computer and decide how you will provide access to it for academics, researchers and members of the public.

**The Details of the Collection:**

The desktop computer is 10 years old, and runs a dated version of Windows.

The computer's hard disk contains a mix of many different kinds of information: the historian's research into medieval war, correspondence with other academics, personal documents, his financial accounts and personal contacts.

There are many versions of the drafts of his major works, saved as they evolved, were edited and improved. The is also a large collection of digital images, both born-digital photographs taken by the historian, as well as digitised material from a variety of sources.

There is a large volume of information. The hard disk has a capacity of 2TB and holds hundreds of thousands of documents and other files.

**Preservation Tool List**

The following is just a small selection of the digital preservation tools that are available. For more information on these tools and many others visit COPTR (Community Owned digital Preservation Tools Registry) <http://coptr.digipres.org/Main_Page>.

**Autopsy (Forensics)** <http://www.sleuthkit.org/autopsy/>

Autopsy is an easy to use, GUI-based program that allows you to efficiently analyze hard drives and smart phones.It is a digital forensics platform and graphical interface to The Sleuth Kit and other digital forensics tools. It is used by law enforcement, military, and corporate examiners to investigate what happened on a computer.

**Bad Peggy (File Validation)** <https://www.coderslagoon.com/index.php?lang=EN>

Bad Peggy scans images ([JPEG](http://fileformats.archiveteam.org/wiki/JPEG), [PNG](http://fileformats.archiveteam.org/wiki/PNG), [BMP](http://fileformats.archiveteam.org/wiki/BMP), [GIF](http://fileformats.archiveteam.org/wiki/GIF)) for damages and other blemishes, and shows the results and files instantly. It enables you to find such broken files quickly, inspect and then either delete or move them to a different location.

**BagIt (Information Package Generator)** <https://github.com/LibraryOfCongress/bagit-java>

The BagIt specification is a hierarchical file packaging format for the creation of standardised digital containers called 'bags,' which are used for storing and transferring digital content. Derived from work by the Library of Congress and the California Digital Library, a bag consists of a ‘payload’ - the digital content - and ‘tags' - metadata files to document the storage and transfer of the bag. There are a number of [Bagit-specific tools](http://sourceforge.net/projects/loc-xferutils/) to ease bag creation, including the BagIt Library, a Java-based software library to support the creation, manipulation, and validation of bags. For those less comfortable with command-line interface, the Bagger application provides a graphical user interface to the BagIt Library.

**C3PO (Collection Profiler)** <http://peshkira.github.io/c3po/>

C3PO helps managers of digital collections get a better understanding of their collection by analyzing data and providing analysis and visualization of the content. C3PO has the ability to aggregate and combine metadata across collections, generate a profile of the content, and use FITS or Apache TIKA metadata. C3PO integrates easily with Plato for preservation planning purposes and with Scout for monitoring preservation risks.

**Dioscuri (Emulator)** <http://dioscuri.sourceforge.net/>

Dioscuri is a Windows-capable x86 hardware emulator for 16-bit operating systems.  Users must supply the original software they wish to run.  It is unable to access physical storage devices, offering instead a virtual floppy drive and virtual hard disk; users must create a disk image from the original physical carrier for any files they wish to access.

**DROID (File Characterisation, Metadata Extractor)** <http://digital-preservation.github.io/droid/>

DROID (Digital Record Object Identification) is a software tool developed to perform automated batch identification of file formats. DROID is designed to meet the fundamental requirement of any digital repository to be able to identify the precise format of all stored digital objects, and to link that identification to a central registry of technical information about that format and its dependencies.

**FITS -File Information Tool Set (File Characterisation, Validator, Metadata Extractor)** <http://projects.iq.harvard.edu/fits>

FITS allows data curators to identify, validate, and extract technical metadata for the objects in their digital repository. It does this by incorporating a range of mostly third-party open source tools, normalising and consolidating their output.

**Fixity (File Fixity Checker)** <https://www.avpreserve.com/avpsresources/tools/>

Fixity is a tool for monitoring file integrity and attendance over time. It is designed to be easy-to-use (GUI interface) and automated (scheduled checks, email alerts/reports) to help small-medium collections ensure the integrity of their digital assets.

**FSlint (De-Duplication Tool)** <http://www.pixelbeat.org/fslint/>

FSlint is a utility to find and clean various forms of unwanted extraneous files on a computer file system. This excess of unnecessary files is referred to as lint. FSlint will help find unwanted or problematic lint in your files or file names. The most common forms of lint FSlint finds are duplicate files, empty directories, and improper names. FSlint has multiple tools to perform a multitude of tasks in both the graphical interface and command line modes.

**Image Magick (Format Migration)** <http://www.imagemagick.org/script/index.php>

ImageMagick is a software suite to create, edit, compose, or convert bitmap images. It can read and write images in a variety of formats (over 100) including DPX, EXR, GIF, JPEG, JPEG-2000, PDF, PNG, Postscript, SVG, and TIFF. Can be used as a migration tool for bitmap images.

**KEEP Emulation Framework (Emulator)** <http://emuframework.sourceforge.net/>

The [KEEP Emulation Framework](http://emuframework.sourceforge.net/) (EF) allows users to view and interact with digital files that otherwise would require obsolete hardware and software.  When confronted with a file, the EF identifies its format and determines what hardware, operating system, and software are needed for proper display. Using a suite of hardware emulators and an archive of software, the EF assembles the proper configuration of components so that users can view the file with its intended ‘look and feel,’ independent of current state-of-the-art computer systems. The EF uses PRONOM and UDFR to identify required software.

**SIARD Suite (Format Migration/Normalisation)** <https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html>

SIARD Suite extracts the contents from a relational database and saves them in the SIARD format which is appropriate for archiving. In this format, the data is stored for future generations and can be uploaded into a new database that may differ from the original. Therefore, this method allows to retain data independently of the original database and to reuse them in the future in modern database systems.

**VeraPDF (Validator)** <http://verapdf.org/>

veraPDF is an open source PDF/A validation tool for Windows, Mac and Linux. The software is still in development, wich is the reason why currently only PDF/A3b, PDF/A2b, PDF/A1b and PDF/A1a can be validated. Results are exported using either the HTML or XML format.

**Xena (Format Migration/Normalisation)** <http://xena.sourceforge.net/>

Xena is free and open source software developed by the National Archives of Australia to aid in the long-term preservation of digital records. Xena is an acronym meaning Xml Electronic Normalising for Archives. When Xena normalises a file, it is converted into a new Xena file with the extension .xena. These files contain the normalised file as well as any extra information relevant to the normalisation process. You can view these files and export the normalised file via the Xena Viewer.

**Preservation Questions**

**Identify Aims and Constraints**

Why do we want to keep the material?

What are we trying to achieve?

For whom are we keeping it?

How do we test their expectations?

What are our constraints in terms of cost/resources?

**Understand the Data, Context and Risks**

What is the collection? How does it break down?

Do we want to retain everything?

Are there any data protection or other legal issues?

What risks do the different parts of the collection face?

What are the highest priorities for action?

**Explore the Options**

What are our preferred preservation approaches? Will we use one approach or a combination?

What actions should we take to achieve them?

What tools do we have available to carry them out?

Will we require additional support to implement these actions?

**Evaluate and Make Decisions**

Can we trial alternative approaches to aid evaluation?

What are our expectations of quality? Do we want to prioritise certain criteria?

How will we validate our plans?

How will we document our decisions?

How and when will we update our plans?

**Exercise Questions**

**What are the main challenges you face?**

**What would success look like?**

**Which tools will you use?**

**What will be your process? (Use separate sheet)**