Technology Watch
Digital Archiving Innovation and Research (DAIR)

• Leading research and innovation in digital archiving policy, methodologies, techniques, and practice
• Promoting and supporting staff digital archiving capability to support the National Archives vision, priorities and strategies
• Contributing subject matter expertise to NAA projects and operations
• Technology watch is a component of our brief
Digital Archives Strategic Research Priorities Framework
Digital Archives Innovation & Research
February 2021
Types of Research

Horizon scanning
- Archival strategies & emerging practice, including indigenous collections
- ‘Stretch’ topics less based in current operations
- May not have a clear implementation pathway
- Often with national / international partners

Commissioned archival practices and techniques
- More attuned to operational needs, pain points and challenges
- Must have an identified business owner within NAA
- Focus on pragmatic outcomes for business improvement
<table>
<thead>
<tr>
<th>Category</th>
<th>Sub Category</th>
<th>Topic/Technology</th>
<th>Description</th>
<th>Sourced from</th>
<th>Date-added</th>
<th>Added by</th>
<th>Referred to</th>
<th>Action Required</th>
<th>Staff member</th>
<th>Completed</th>
<th>Date last updated</th>
<th>Updated by</th>
<th>Outcome</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>digital preservation</td>
<td>Emulation as a Service Infrastructure</td>
<td>A project to develop a scalable emulation to support a shared emulation capability</td>
<td>Yale University, Software Preservation Network</td>
<td>6-Jan-21</td>
<td>JD</td>
<td>Test</td>
<td>James Dog, Carey Garvie, Tim Mitsud, Bridget Dexter</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2019/3187</td>
</tr>
<tr>
<td>Technology</td>
<td>digital preservation</td>
<td>Virtual Research Environment (VRE) - DDAM &amp; QPS</td>
<td>virtual platform containing key digital preservation tools</td>
<td>DP JISC listserv</td>
<td>5-Jan-21</td>
<td>CG</td>
<td>Test</td>
<td>Yaso Arumugam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2020/4159</td>
</tr>
<tr>
<td>Technology</td>
<td>email</td>
<td>Review, Appraisal and Triage of Mail (RATOM) - UNC, Chapel Hill</td>
<td>software to assist with email analysis, selection and appraisal tasks</td>
<td></td>
<td>5-Jan-21</td>
<td>CG</td>
<td>Test</td>
<td>James Dog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>digital forensics</td>
<td>BitCurator environment - UNC, Chapel Hill</td>
<td>suite of digital forensics tools</td>
<td>Cal Lee</td>
<td>5-Jan-21</td>
<td>CG</td>
<td>Test</td>
<td>James Dog, Carey Garvie</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>risk management</td>
<td>Digital Preservation Framework - NARA</td>
<td>NARA risk management approach to digital preservation of file formats. Includes risk matrix and preservation action plans for over 300 file formats</td>
<td>NARA</td>
<td>5-Jan-21</td>
<td>CG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2020/984</td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>risk management</td>
<td>DiARARM - Digital Archives Graphical Risk Assessment Model - TNA &amp; Unit of Warwick</td>
<td>Tool to assess level of risk to digital collections utilising Bayesian network</td>
<td>DPC listserv</td>
<td>5-Jan-21</td>
<td>CG</td>
<td>Watch</td>
<td>James Dog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2020/2173</td>
</tr>
<tr>
<td>Technology</td>
<td>file format</td>
<td>PRONOM - TNA</td>
<td>File Format Registry developed by TNA (supports DROID - file format identification tool), TNA proposing key updates as outlined in RM102021 - Includes new data model (graph data, linked data)</td>
<td>OFF Conference documentation 2020</td>
<td>5-Jan-21</td>
<td>CG</td>
<td>Watch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2021/76</td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>web archiving</td>
<td>Web Curator Tool (WCT) - NLNZ &amp; NL, Netherlands</td>
<td>open source workflow management tool for selective web harvesting. It supports selecting, crawling websites, performing QA and preparing web sites for ingest to archival storage.</td>
<td>OFF Conference documentation 2020</td>
<td>5-Jan-21</td>
<td>CG</td>
<td>Watch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2021/76</td>
<td>Web Archiving Tool, may be useful if NAA decides to undertake this work in future</td>
</tr>
</tbody>
</table>
Product Assessment Template

**Product: Libreron**

**Assessment Date:** 19/01/2021

**Assessor(s):** James Doe

**Recommendation:**

That the Archives continue to test Libreron with a view to developing a pre-import/guest process for PST files that incorporates the product (e.g. a simplified report, or the full SQLite file, could be ingested with the original PST file).

**Assessment:**

This product has a practical use case for the Archives, is easy to install and deploy, and is well-supported by a community of archivists, libraries, and universities. The product can run on Ubuntu, MacOS, and Windows 10. It has been tested only on Ubuntu 20.04 LTS.

**Product Overview:**

Libreron is a product of the RATOM (Review, Appraisal, and Traceability of Metadata Project) led by the University of North Carolina in the US. Libreron consists of a Python library and supporting utilities to parse and process PST (Microsoft Outlook Personal Folder) and MBOX (MBOX Email Format) email files. Libreron is a command line-based tool.

Libreron extracts data from the PST file into a SQLite file which the user then imports into a SQLite database. The data can be viewed and interrogated via the SQLite database.

The application can run on Ubuntu, MacOS, and Windows 10.

The available commands include:

- **Entity extraction** extracts data from individual PST and mbox files, or directories containing PST and mbox files. Message files, date, full message content can be seen and interrogated. Attachments are extracted and listed and linked to the source email.

- **Scan and report** quickly scans an email source and generates a report (file metadata, message content, and attachment metadata).

- **Message export** exports selected messages from PST files as one (and Electronic mail format for individual emails) file per message.

**Use Cases:**

Libreron can provide detailed reports on PST exports of Outlook email systems that are transferred into the Archives' custody, for example, Mr Perry's Emails (CO ROM - In-Office and Outlook). Item barcode 32710381. This item consists of a single PST file stored on a CD-ROM, i.e., it has not been processed into the digital archive. There is no metadata about the record(s) or index of the contents of the PST file. Libreron would...
Communication Plan

Why? DAIR goals

Who? Audience

What? Message

How? Channels
Thank You!