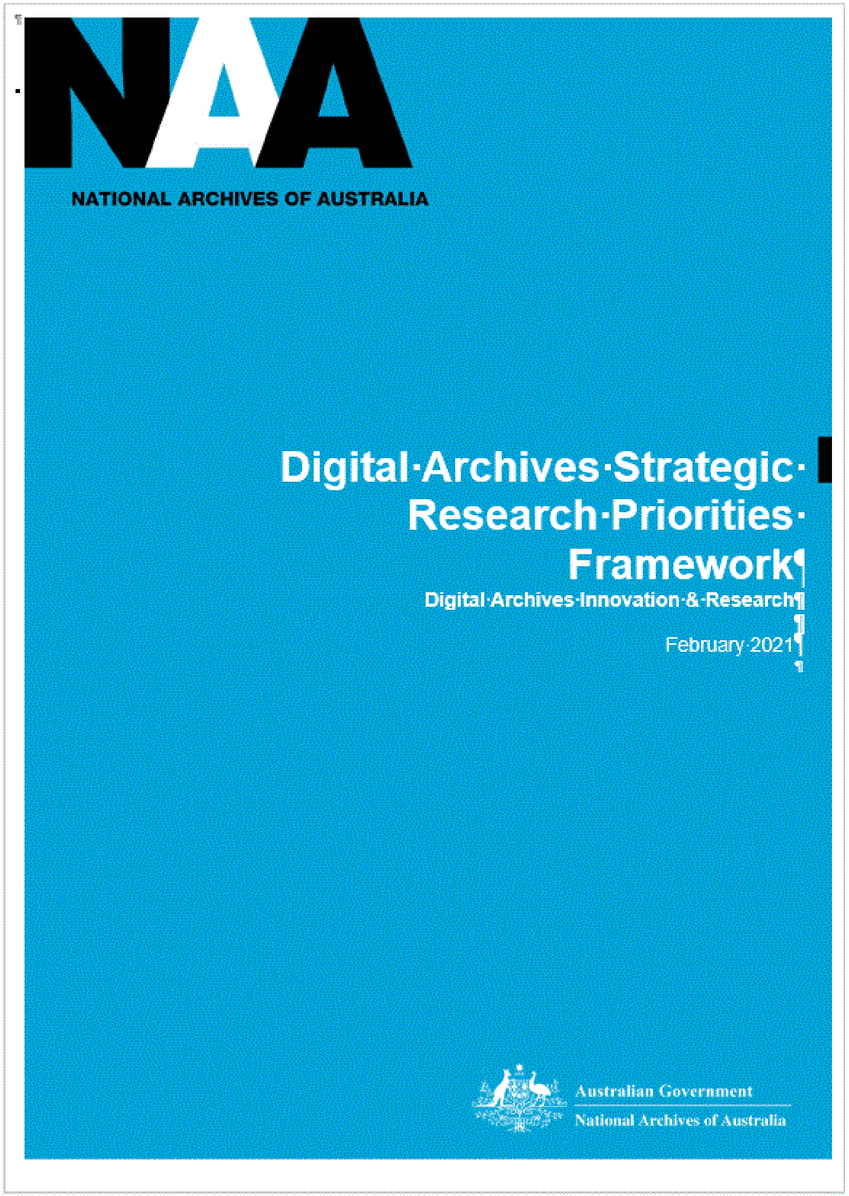
NATIONAL ARCHIVES OF AUSTRALIA

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



NAA

NATIONAL ARCHIVES OF AUSTRALIA

Digital Archives Strategic Research Priorities Framework

Digital Archives Innovation & Research

February 2021

Australian Government

National Archives of Australia

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

dpc

DigitalPreservationCoalition

E-ARK

DCC

ICA

International Council on Archives

Conseil International des Archives

Software Preservation Network

Open Preservation Foundation

pasig

preserving and archiving special interest group

CAARA

Council of Australasian Archives and Records Authorities

WE MISS iPRES

Jisc

NDSA

International Association of Sound and Audiovisual Archives

Internationale Vereinigung der Schall- und audiovisuellen Archive

Association Internationale d'Archives Sonores et Audiovisuelles

Asociación Internacional de Archivos Sonoros y Audiovisuales

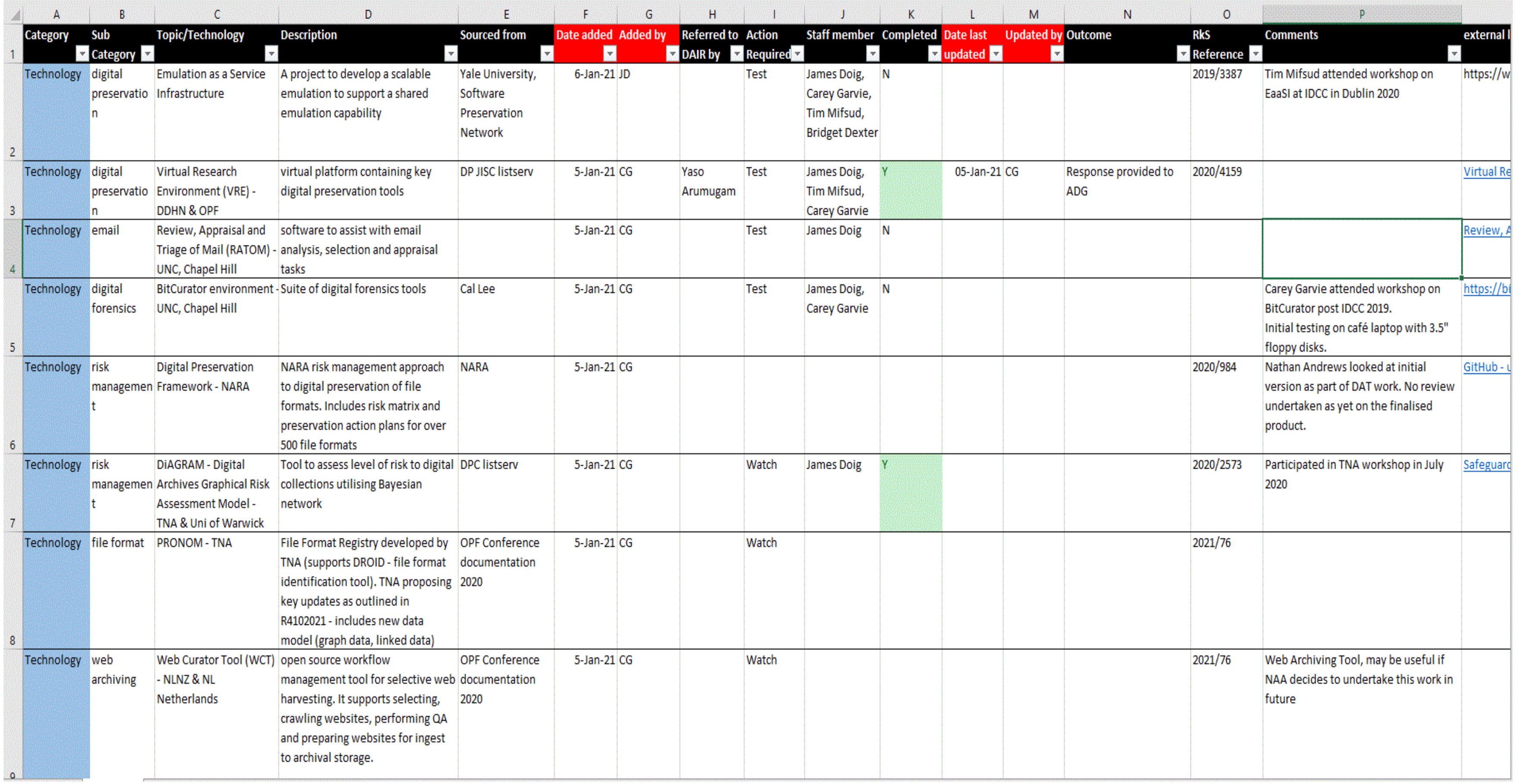
iasa

FADGI

Federal Agencies Digital Guidelines Initiative

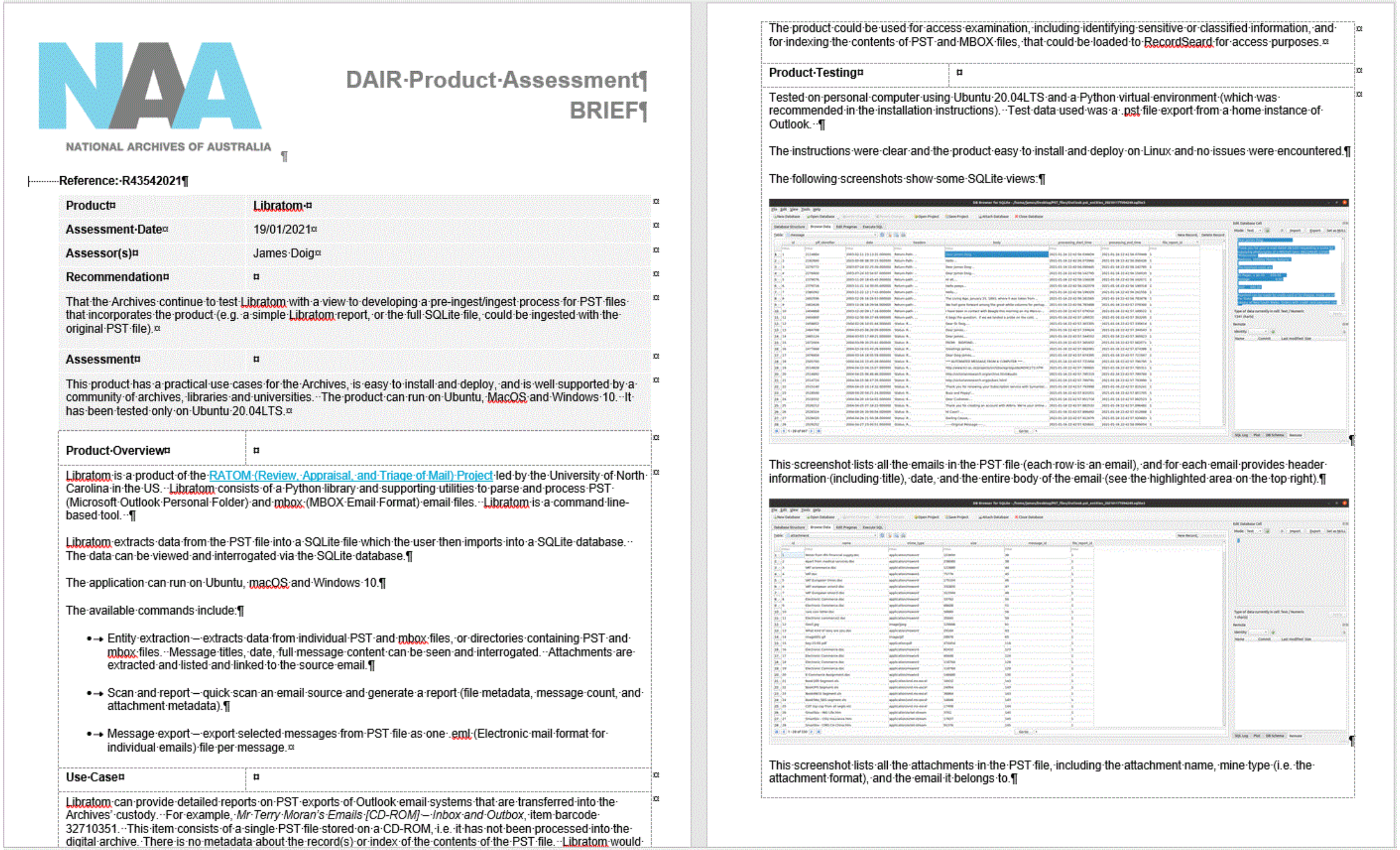
dce

Horizon Scanning and Technology Watch Register



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Category	Sub Category	Topic/Technology	Description	Sourced from	Date added	Added by	Referred to DAIR by	Action Required	Staff member	Completed	Date last updated	Updated by	Outcome	RkS Reference	Comments	external l
2	Technology	digital preservation	Emulation as a Service Infrastructure	A project to develop a scalable emulation to support a shared emulation capability	Yale University, Software Preservation Network	6-Jan-21	JD		Test	James Doig, Carey Garvie, Tim Mifsud, Bridget Dexter	N				2019/3387	Tim Mifsud attended workshop on EaaSI at IDCC in Dublin 2020	https://w
3	Technology	digital preservation	Virtual Research Environment (VRE) - DDHN & OPF	virtual platform containing key digital preservation tools	DP JISC listserv	5-Jan-21	CG	Yaso Arumugam	Test	James Doig, Tim Mifsud, Carey Garvie	Y	05-Jan-21	CG	Response provided to ADG	2020/4159		Virtual Re
4	Technology	email	Review, Appraisal and Triage of Mail (RATOM) - UNC, Chapel Hill	software to assist with email analysis, selection and appraisal tasks		5-Jan-21	CG		Test	James Doig	N						Review, A
5	Technology	digital forensics	BitCurator environment - UNC, Chapel Hill	Suite of digital forensics tools	Cal Lee	5-Jan-21	CG		Test	James Doig, Carey Garvie	N					Carey Garvie attended workshop on BitCurator post IDCC 2019. Initial testing on café laptop with 3.5" floppy disks.	https://bi
6	Technology	risk management	Digital Preservation Framework - NARA	NARA risk management approach to digital preservation of file formats. Includes risk matrix and preservation action plans for over 500 file formats	NARA	5-Jan-21	CG								2020/984	Nathan Andrews looked at initial version as part of DAT work. No review undertaken as yet on the finalised product.	GitHub - u
7	Technology	risk management	DiAGRAM - Digital Archives Graphical Risk Assessment Model - TNA & Uni of Warwick	Tool to assess level of risk to digital collections utilising Bayesian network	DPC listserv	5-Jan-21	CG		Watch	James Doig	Y				2020/2573	Participated in TNA workshop in July 2020	Safeguard
8	Technology	file format	PRONOM - TNA	File Format Registry developed by TNA (supports DROID - file format identification tool). TNA proposing key updates as outlined in R4102021 - includes new data model (graph data, linked data)	OPF Conference documentation 2020	5-Jan-21	CG		Watch						2021/76		
9	Technology	web archiving	Web Curator Tool (WCT) - NLNZ & NL Netherlands	open source workflow management tool for selective web harvesting. It supports selecting, crawling websites, performing QA and preparing websites for ingest to archival storage.	OPF Conference documentation 2020	5-Jan-21	CG		Watch						2021/76	Web Archiving Tool, may be useful if NAA decides to undertake this work in future	

Product Assessment Template



NATIONAL ARCHIVES OF AUSTRALIA

DAIR Product Assessment
BRIEF

Reference: R43542021

Product	**Libratom**
Assessment Date	19/01/2021
Assessor(s)	James Doig
Recommendation	

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

Assessment

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

Product Overview

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

Libratom 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 titles, date, full message content can be seen and interrogated. Attachments are extracted and listed and linked to the source email.
- Scan and report – quick scan an email source and generate a report (file metadata, message count, and attachment metadata).
- Message export – export selected messages from PST file as one .eml (Electronic mail format for individual emails) file per message.

Use Case

Libratom can provide detailed reports on PST exports of Outlook email systems that are transferred into the Archives' custody. For example, *Mr Terry Moran's Emails [CD-ROM] – Inbox and Outbox*, item barcode 32710351. 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. Libratom would

The product could be used for access examination, including identifying sensitive or classified information, and for indexing the contents of PST and MBOX files, that could be loaded to RecordSearch for access purposes.

Product Testing

Tested on personal computer using Ubuntu 20.04LTS and a Python virtual environment (which was recommended in the installation instructions). Test data used was a .pst file export from a home instance of Outlook.

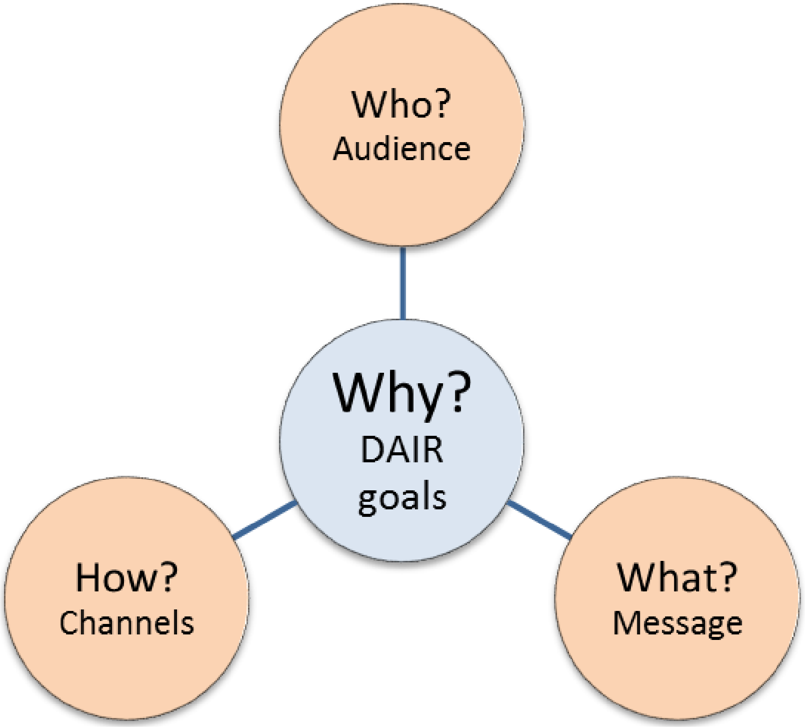
The instructions were clear and the product easy to install and deploy on Linux and no issues were encountered.

The following screenshots show some SQLite views:

This screenshot lists all the emails in the PST file (each row is an email), and for each email provides header information (including title), date, and the entire body of the email (see the highlighted area on the top right).

This screenshot lists all the attachments in the PST file, including the attachment name, mine type (i.e. the attachment format), and the email it belongs to.

Communication Plan



Thank You!