

Email Archives at the University of Manchester

Preservation and management projects with ePADD

Jessica Smith, Paul Carlyle & Jochen Farwer

Records Management und Digitale Archivierung, Universität Graz, May 2023

Carcenet Press Email Archive

- ❖ Accessions 1 & 2 – 300, 000 items (email and attachments)
- ❖ Accession 3 – approx. 482,000 items (email and attachments)

Current total figures from ePADD:

- Messages: **447,848**
- Attachments: **270,414**
- Images: **174,999**
- Documents: **88,167**
- Others: **7,248**

Preservation

- ❖ Preservica



Access

- ❖ ePADD





Developed by Stanford University Libraries

- ❖ Labels
- ❖ Natural language processing
- ❖ Entity identification
- ❖ Lexicon analysis

ePADD+ Project

January 2021 - December 2022

Email represents an increasingly important part of the historical record. Preserving and ensuring access to this record are therefore central to the functions and values of archives and archivists. Until we can meet the challenges of email archiving, responsible custody is undermined, accountability is abandoned, and, ultimately, the historical record is imperiled. In short, the problem won't take care of itself, and the time to act is now.

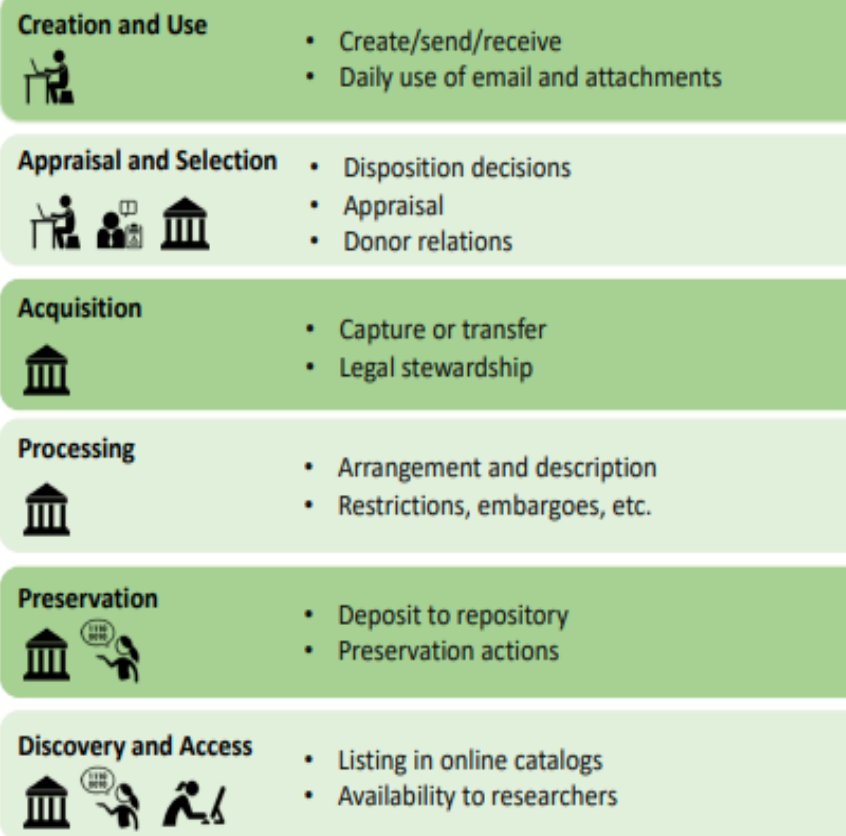
The Future of Email Archives (2018), p. 3, www.clir.org/pubs/reports/pub175/



Applied to email records

- Acquire, as files from email clients/servers or IMAP protocol
- Save original order, e.g., headers, multi-parts, attachments
- Characterize and describe
- Normalize/create derivatives
- Persistent protection and security
- Discovery/delivery

E@S & ePADD
Functional overlap
and added value



ePADD

E@S,
ePADD

E@S

ePADD

ePADD+ Project



- Integrate EAS-like long-term preservation function with ePADD's existing support for appraisal, processing, discovery, and delivery
- Collaborative effort by Harvard, Manchester, Stanford, and Amuse Labs
- Funded by University of Illinois (#098468-18280) as part of its EA:BCC regrant initiative, itself supported by the Andrew W. Mellon Foundation (#1905-06768)

New Preservation Features in ePADD

ePADD Versions 9.0 and 10.0



New features: Import

Previous ePADD versions



After ePADD+ Project

- Limited header information was captured (To, from, subject, date...)
- Some attachments not handled (e.g. text, html)
- Email text captured but formatting was lost
- No support for multi-part messages

- Complete header information stored
- All attachments types are handled
- Message formatting is captured
- Multi-part messages are handled
- Optional format derivatives
- Optional 'sidecar' files



New features: Metadata

Previous ePADD versions



After ePADD+ Project

- Limited metadata captured
- Stored in multiple places

- Much richer set of metadata captured:
 - Automatic metadata (e.g. PREMIS events)
 - Manual metadata (additional fields available in GUI)
- Metadata stored in a single place using PREMIS



New features: Export

Previous ePADD versions



After ePADD+ Project

- Appraisal module export included only post-appraisal data for processing module
- Key preservation info e.g. MBOX structure, duplicates/folder names etc. not included

- 'Preservation ready' Bag structured to include multiple exportable assets including:
 - Full MBOX structure
 - Canonical version of archive
 - Derivative versions
 - Additional sidecar files
 - Metadata

New features: Emailchemy integration

Previous ePADD versions



After ePADD+ Project

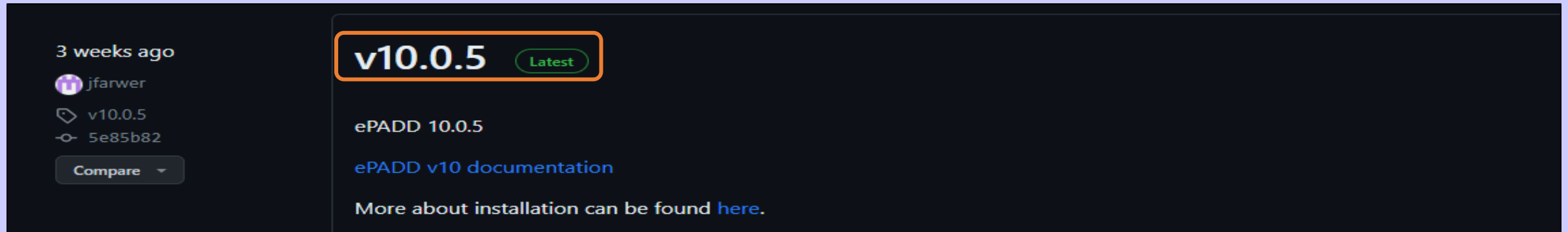
- Conversion to mbox format required outside of ePADD application

- On purchase of an Emailchemy licence and creation of dongle
- Conversion to mbox from multiple email formats can take now place within ePADD
- Conversion of appraised or processed files to EML format is also possible on export

How to download ePADD

You can download ePADD 10.0 from GitHub:

<https://github.com/ePADD/epadd/releases>



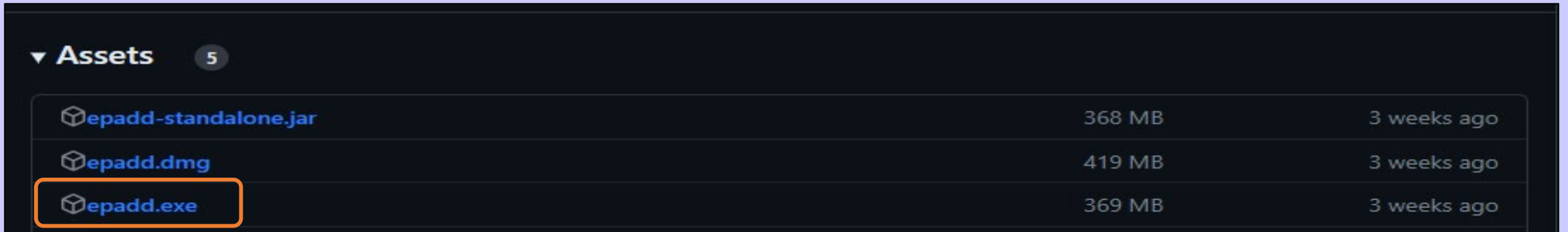
3 weeks ago
jfarwer
v10.0.5
5e85b82
Compare

v10.0.5 Latest

ePADD 10.0.5

[ePADD v10 documentation](#)

More about installation can be found [here](#).



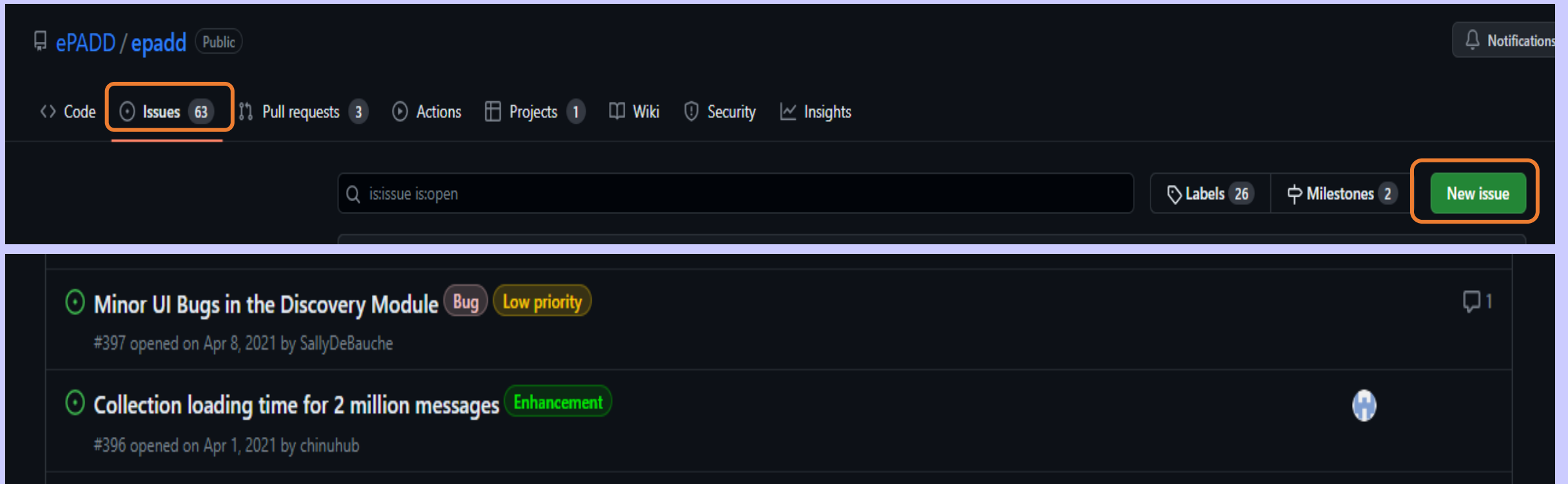
Assets 5

epadd-standalone.jar	368 MB	3 weeks ago
epadd.dmg	419 MB	3 weeks ago
epadd.exe	369 MB	3 weeks ago

Reporting issues

Report issues or request enhancements on the GitHub page:

<https://github.com/ePADD/epadd/issues>



ePADD / epadd Public

Code **Issues 63** Pull requests 3 Actions Projects 1 Wiki Security Insights

is:issue is:open Labels 26 Milestones 2 **New issue**

- Minor UI Bugs in the Discovery Module **Bug** **Low priority** #397 opened on Apr 8, 2021 by SallyDeBauche
- Collection loading time for 2 million messages **Enhancement** #396 opened on Apr 1, 2021 by chinuhub

New website



The screenshot shows the homepage of the ePADD website. At the top, there is a navigation bar with the ePADD logo and the text "Email Archiving Tool". The navigation menu includes "Home", "About", "Using ePADD", "Resources", and "Community". The main header features the text "ePADD" in large white letters on a dark blue background with a pattern of envelope icons. Below this, the full name "Email Processing, Appraisal, Discovery, Delivery (and Preservation!)" is written in white. A paragraph of text describes the tool as free and open source software developed by Stanford University's Special Collections & University Archives, supporting the appraisal, processing, preservation, discovery, and delivery of historical email archives. Below the text are four icons with corresponding links: "Download and Install" (ePADD logo), "User Guide" (book icon), "Emailchemy for ePADD" (purple cone with @ symbol), and "Get Involved" (network of people icon). At the bottom, there is a "Contact ePADD Project" button and social media icons for GitHub, YouTube, and Twitter.

Email Archiving Tool

Home About Using ePADD Resources Community

ePADD

Email Processing, Appraisal, Discovery, Delivery (and Preservation!)

ePADD is free and open source software developed by Stanford University's Special Collections & University Archives that supports the appraisal, processing, preservation, discovery, and delivery of historical email archives. ePADD incorporates techniques from computer science and computational linguistics, including machine learning, natural language processing, and named entity recognition to help users access and search email collections of historical and cultural value.

[Download and Install](#)

[User Guide](#)

[Emailchemy for ePADD](#)

[Get Involved](#)

Contact ePADD Project



Steering Group

Join the Steering Group to:

- Advance the ePADD roadmap through collective efforts and strategic grant applications
- Collaborate with other ePADD users on improving documentation and enhancing user guidance
- Share ideas on workflows, projects, and more

Quarterly open meetings:

- Tuesday, 11 July 2023, 11am - 12pm ET
- Tuesday, 12 September 2023, 11am - 12pm ET
- Tuesday, 14 November 2023, 11am - 12pm ET



Code Group

Join the Code Group to:

- Learn how to contribute to the ePADD codebase
- Help address maintenance and security issues with the tool
- Merge your enhancements in a sustainable way
- Improve code documentation and testing

Monthly open meetings:

- Thursday, 8 June 2023, 10am-11am 2023, ET
- Thursday, 6 July 2023, 10am-11am ET
- Thursday, 14 September 2023, 10am-11am ET
- Thursday, 12 October 2023, 10am-11am ET
- Thursday, 9 November 2023, 10am-11am ET
- Thursday, 7 December 2023, 10am-11am ET



Join the ePADD
community!

Stay informed

- Website
- ePADD listserv
- Twitter: @e_padd

Get involved

- ePADD steering group
- ePADD code group
- ePADD user forum

Project Outputs

- End-to-end workflows and tools for the ingest of email archives
- Introduce digital tools for providing mediated access to the email archives, including ePADD, and integration with digital scholarship tools
- Guidelines for archivists and curators on managing and facilitating access to email archives
- Guidance for writers and small publishers on managing and preserving their emails
- Workshops to seek input from colleagues in other institutions involved in managing born-digital archives
- Creative responses to the email archive

Palladium Project team:

Paul Carlyle, project archivist

Jessica Smith, project manager

Ian Gifford, head of Digital Development

Jochen Farwer & Shiu hung Au yeung, developers



Appraisal

Funnel approach

- Employment/HR
- Office Administration
- Financial
- Facilities
- Student Records
- Personal
- Ephemera

Appraisal Rubric (sample section)

Records which will be retained	Records which will not be retained
Publishing/Literary 1. Correspondence with authors (a) Editorial (b) Readings (c) Marketing (d) Design (e) Contracts/agreements 1. Design 2. Royalties 3. Permissions 4. PN Review	
Employment/HR 1. Employees' handbook	Employment/HR 1. Personal Development Reviews (PDR) 2. Recruitment 3. References (includes requests submitted to Carcanet from individuals or on their behalf) 4. Work experience 5. Apprenticeships 6. Volunteers/Interns 7. Pensions 8. Payroll 9. Employment contracts 10. Grievance/disciplinary proceedings
	Office Administration 1. Office hours 2. Leave/holiday 3. Lottery syndicate 4. Automatic replies/out-of-office replies

Appraisal – Capstone approach

Collaborative appraisal

Alison Brackenbury	504
John Gallas	725
Rory Waterman	705

Total: 1374 email

Project archivist appraisal – Elaine Feinstein

Email: 1526 email

Number restricted: 231

Number cleared for transfer: 1295

Dates: 2002 - 2019

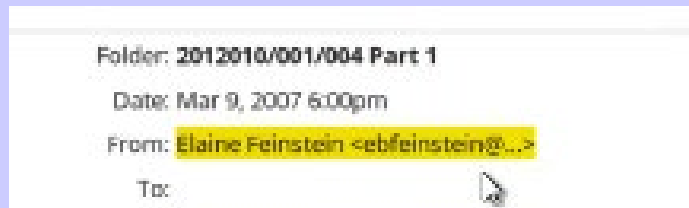
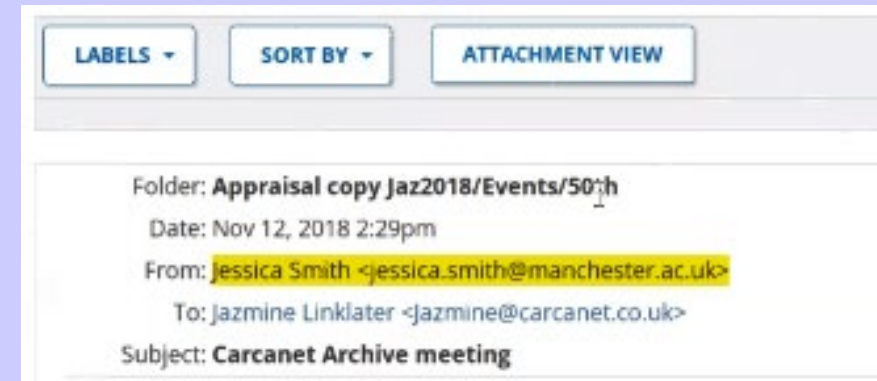
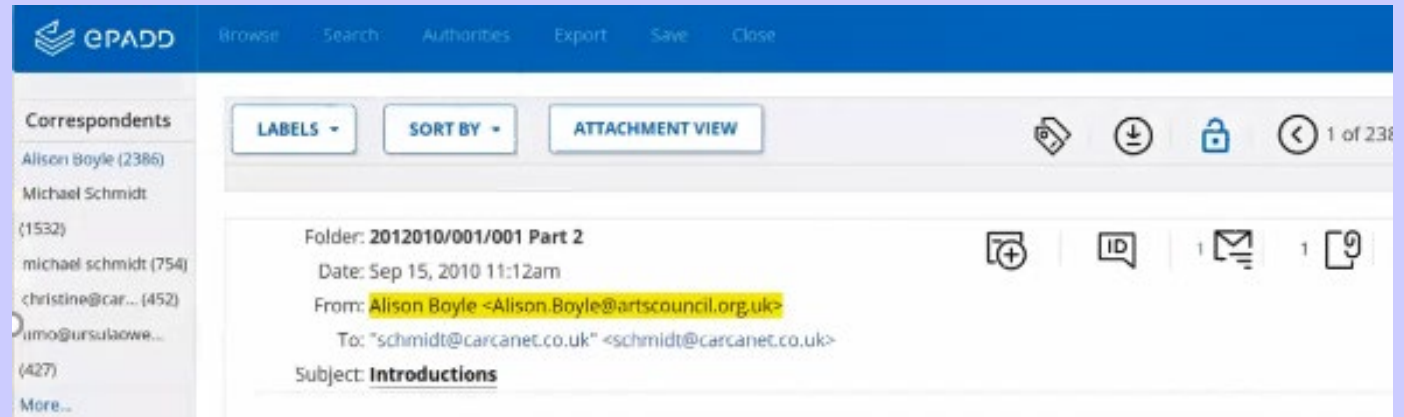
Time taken: estimated 2 weeks

Categories identified for restriction:

- Medical information
- Information about family members
- Personal information about individuals still living
- Contact information for individuals still living
- Private or group discussions of the judging panels of poetry competitions

ePADD Development Work

- .csv download
- Visible folder structure
- Correspondent editing
- Transfer of selected messages to Delivery module
- Delivery module display
- Redaction of email text



New functionality - redaction

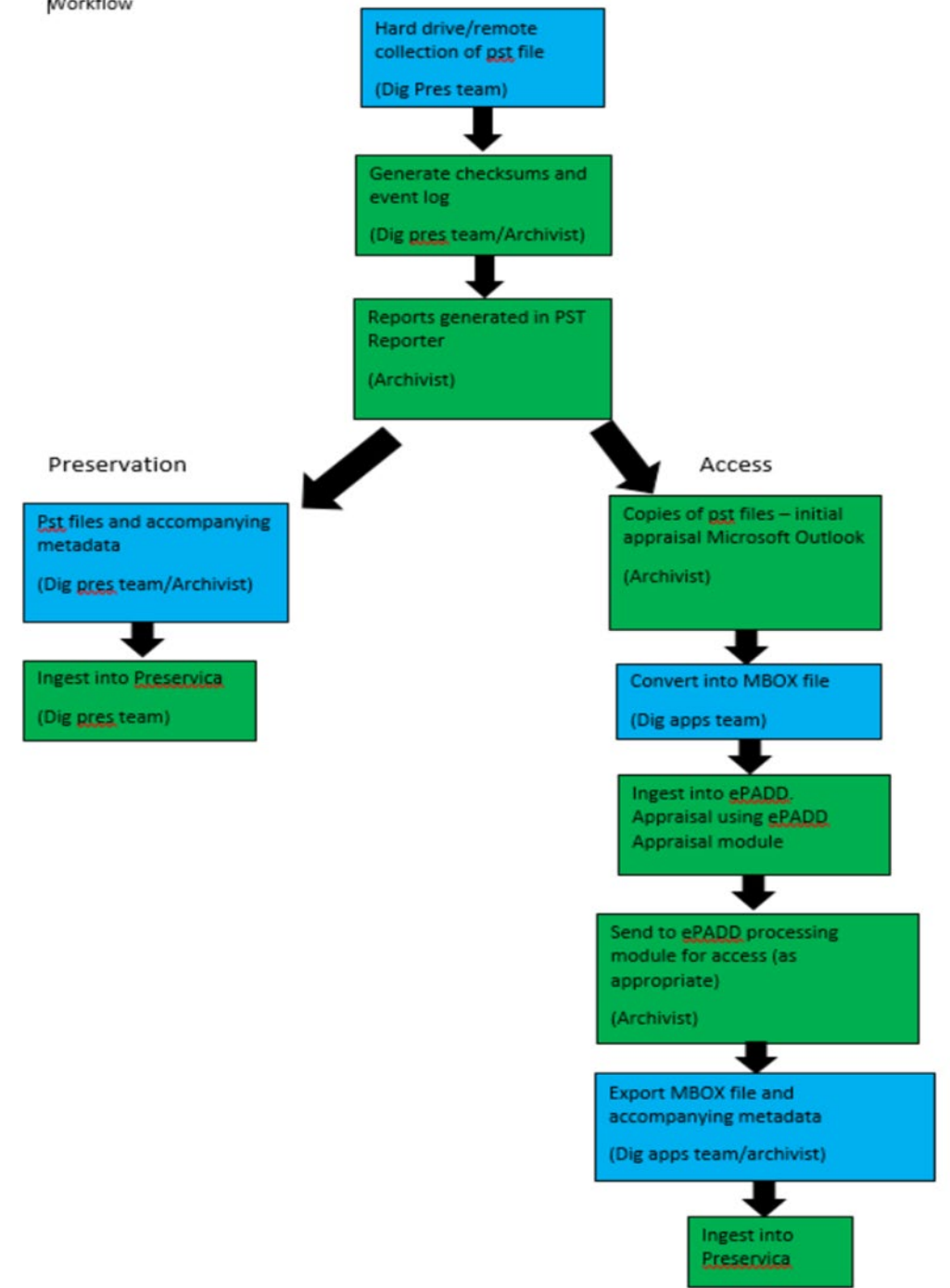




Project Outputs - Documentation

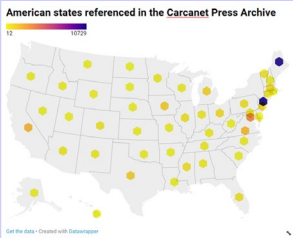
- ❖ Workflow
- ❖ Project Report
- ❖ Guide for archivists and ePADD users
- ❖ Guide for collection donors and writers
- ❖ Project promotion

Workflow

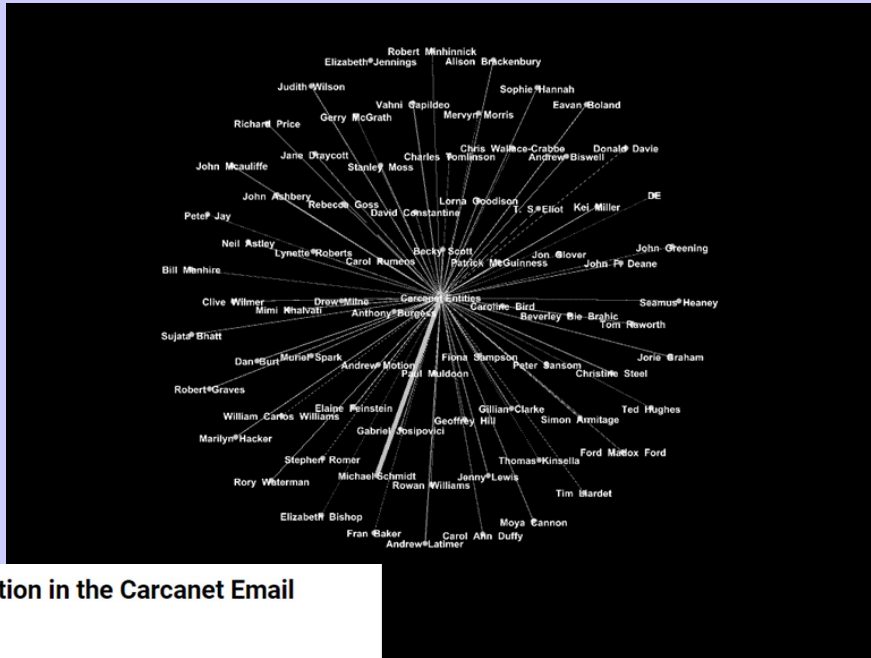


Data Visualisations

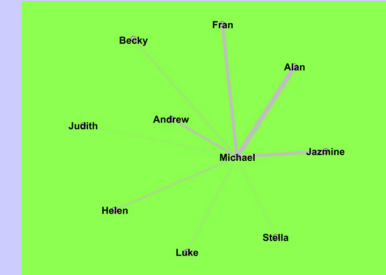
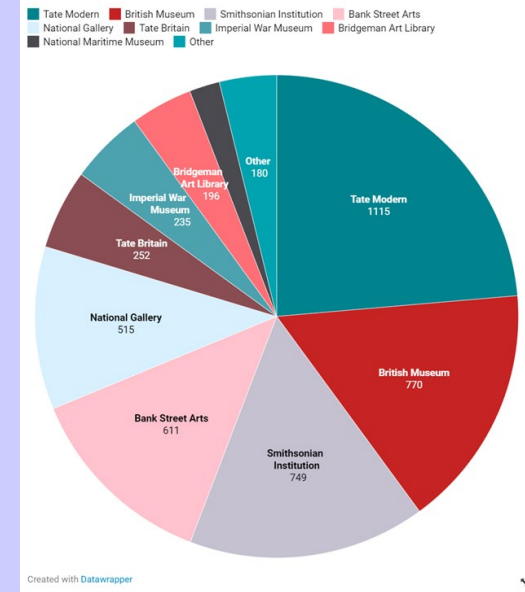
Created in Gephi & Datawrapper, data manually edited



Correspondents in the Carcanet Email Archive



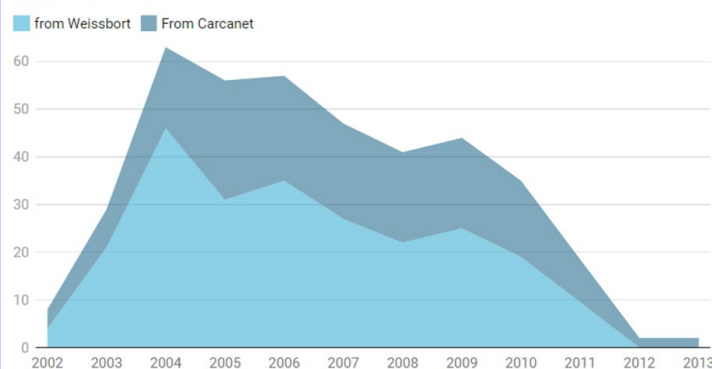
Museums and Galleries in the Carcanet Press Email Archive



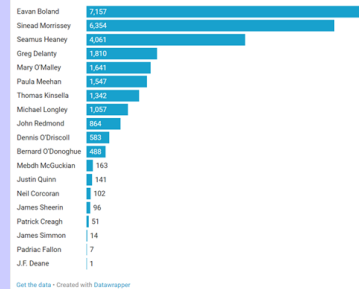
Geographical distribution of poets published by Carcanet 1969 - 2019



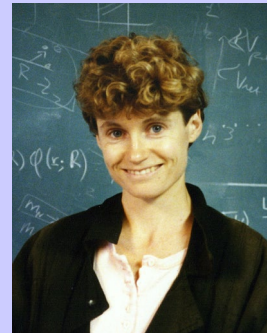
Daniel Weissbord and translation in the Carcanet Email Archive



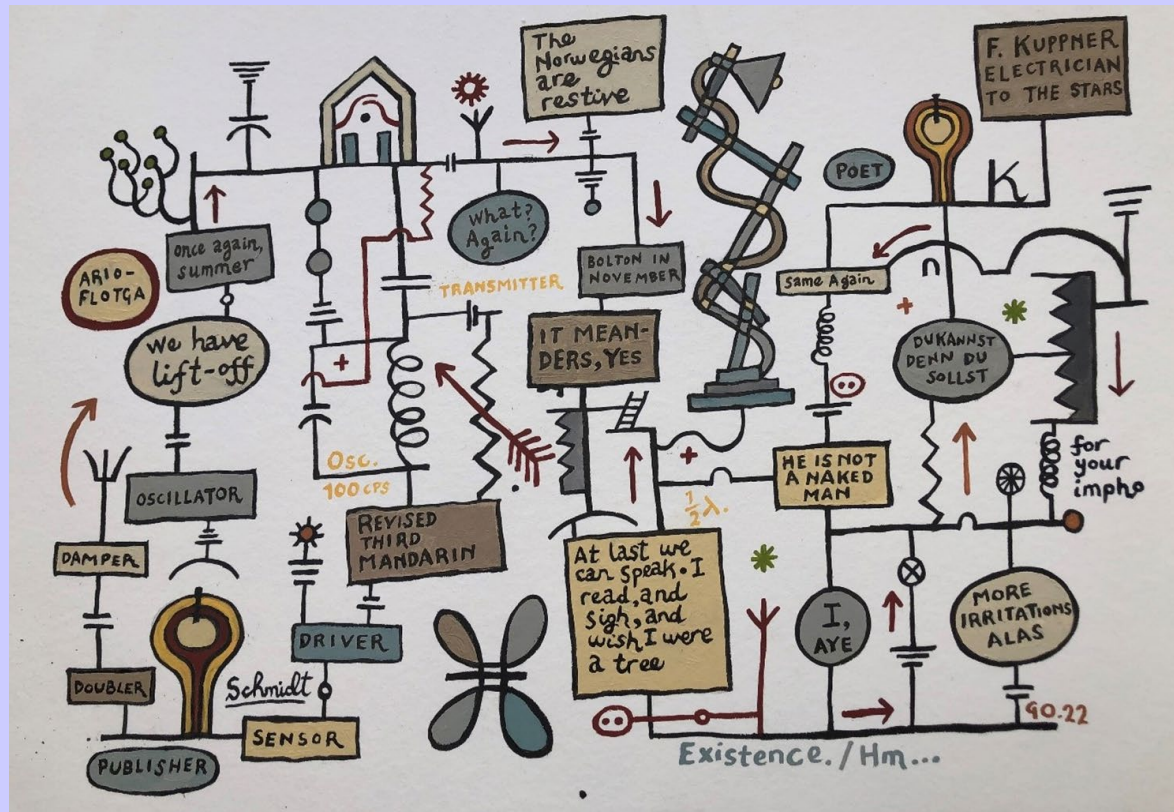
Irish Poets in the Carcanet Press Email Archive



Creative Responses



Rebecca Elson

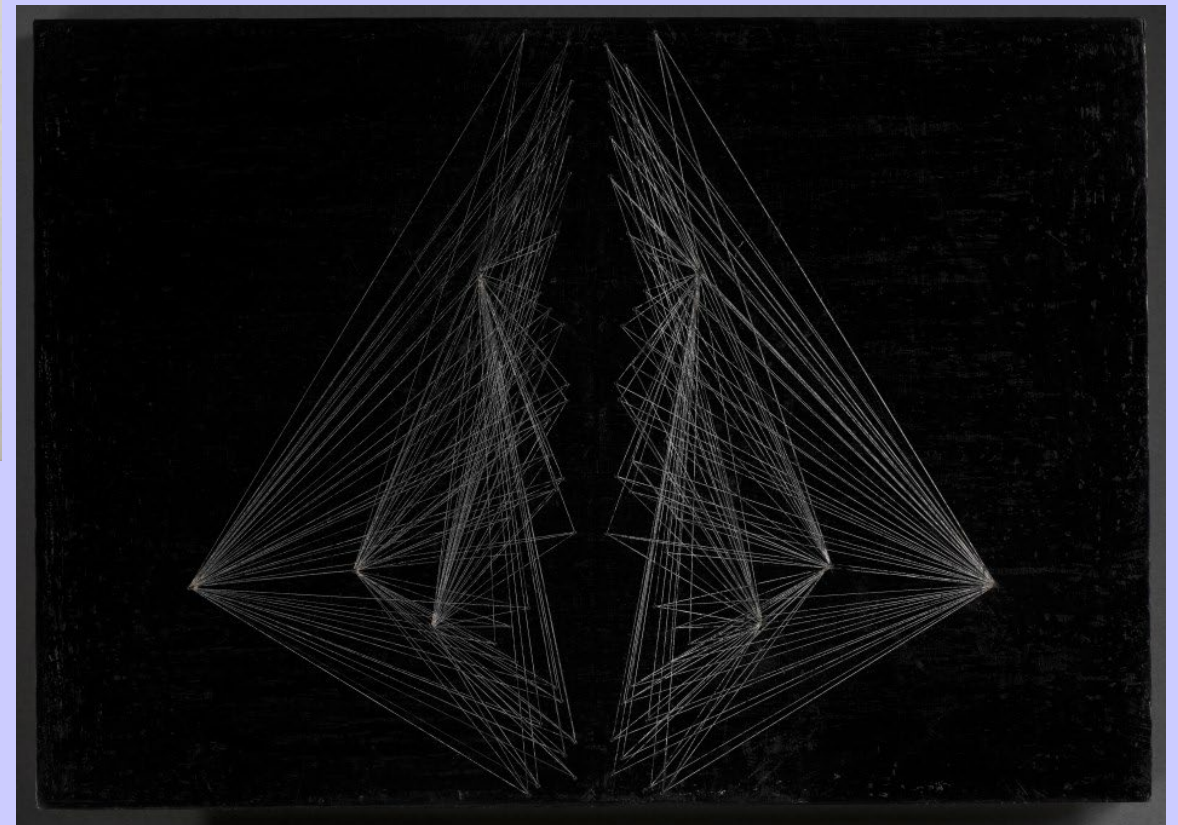


A circuit diagram for Frank Kuppner and Michael Schmidt, Gregory O'Brien



Frank Kuppner

For this we go out dark nights, Mary Griffiths



Conclusions

Appraisal

- A more stringent appraisal rubric
- Bulk appraisal
- Transferring fewer email from the donor
- Improvement of Artificial Intelligence appraisal tools
- New digital records appraisal policy
- Revised Access Permissions in Donation agreements
- Collaborative appraisal

Collaboration is crucial!

Thank you!

Jessica.Smith@manchester.ac.uk

Twitter: @JessicaSmithCAA

Rylands Blog - <https://rylandscollections.com/>