Case Study: The National Library of New Zealand's experiments with ePADD using the Ian Wedde Email Archives

Kaila Fewster[1]

Educational applications: This case study explores the challenges of processing email archival collections to improve access and discoverability, and highlights the potential applications of AI tools to providing solutions to these problems. This case illustrates the challenges of appraising born-digital collections like email archives, in particular in regards to privacy and redaction, and demonstrates how AI tools have potentialities and limitations to integrate into existing traditional archival workflows.

Educational topics: AI for born-digital records, evaluating AI/ML models for archives, privacy concerns using AI, AI for description and access[2].

About: This case study is part of a series of learning materials developed by InterPARES Trust AI[3] researchers and educators to train archival professionals and students to effectively leverage artificial intelligence in their archival work. The final draft was completed on November 19th, 2023. It has a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International BY-NC-SA 4.0 license, which requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.[4]

This case study describes a 2016 collaboration between several colleagues at the National Library of New Zealand Te Puna Mātaurangao Aotearoa to use ePADD v.2 to process the email archives of New Zealand poet Ian Wedde. The National Library had accessioned a number of born-digital archives, but the significant labour involved in appraisal and potential privacy implications for the donor meant that these collections were generally not made available to researchers. At this point, the appraisal process for email accounts involved manually identifying and redacting sensitive and personal information, including financial details, government identification, and restricted topics from a list provided by the donor. To assist with content appraisal and accessibility of these email collections, the library became an early adopter of ePADD, an open-source software developed by Stanford Libraries for appraising, processing, and providing access to email archives. The project team received permission from Ian Wedde, a

---

donor and local poet, to use his archived email inbox to test ePADD's functionality with content appraisal and facilitating access.

While developing the Ian Wedde email archives project as a test case, ePADD was particularly helpful in the entity extraction process to get a sense of the content and scope of the emails. Rather than a human manually combing through the full account to get a sense of what was there, ePADD was able to spit out a list of all correspondents and topics, as well as a list of emails containing each entity. Using this information, archivists were able to describe the email collection much more easily. The library has continued to use ePADD occasionally to process its email collections and the software has primarily been utilized to gain a better sense of collections as part of a broader arrangement and description workflow. The library has not yet implemented ePADD as an access tool in the reading room, but that is a possibility they will explore that avenue in the future. Beyond access and discovery, many archival workflows were not adequately addressed by the platform. First, a format migration from the email inbox's original state needed to be performed before processing in ePADD could occur. There was also limited inter-system compatibility between the email account in ePADD and the larger digital preservation system used by the library. In all, this example demonstrates that AI tools like ePADD can act as pathways into digital material, but do not serve as a complete replacement for traditional arrangement and description methods.

Potential Discussion Questions:
1. What are the advantages and limitations of using AI tools for traditional archival processes like content appraisal, arrangement, and description?
2. Based on this case study, what is the relationship between AI tools and technologies and traditional archival methods? In what situations might traditional (or manual) methods still be preferred over AI tools?
3. How might the use of AI tools impact the role and responsibilities of archivists and records managers, and what new skills or knowledge might they need to effectively use these tools?
4. How can archivists and records managers measure the effectiveness and efficiency of AI tools like ePADD in archival processing work? What metrics or indicators should be used to evaluate their performance?

References

Schneider, J. (n.d.). *Josh Schneider | Libraries*. Retrieved November 18, 2023, from
  https://library.stanford.edu/people/josh-schneider

Schneider, R., Adams, C., DeBauche, S., Echols, R., McKean, C., Moran, J., & Waugh, D. (2019*).
  Appraising, processing, and providing access to email in contemporary literary archives*,
  Archives and Manuscripts, 47:3, 305-326, DOI: 10.1080/01576895.2019.1622138