Dr Lise Jaillant

l.jaillant@lboro.ac.uk

Twitter: @lisejaillant

THE FUTURE OF ACCESS TO DIGITAL RECORDS: A USER'S PERSPECTIVE

LUSTRE Workshop 4

Loughborough University

University*of*Reading

THE UNIVERSITY OF BRITISH COLUMBIA

Lise Jaillant, Ph.D. candidate
University of British Columbia
Department of English
#397 – 1873 East Mall
Vancouver, B.C. V6T 1Z1
Canada

Email: ljaill01@interchange.ubc.ca

10 March 2012

Seán Sweeney, Trustee of the estate of James Joyce
Ráith House
Newport
Co. Mayo
Ireland

14 February 2012

Dear Mr. Sweeney,

Dear Ms. Jaillant, We are unaware of the necessity to grant permission for consultation of documents. Thank you for your letter which was forwarded from Ireland.
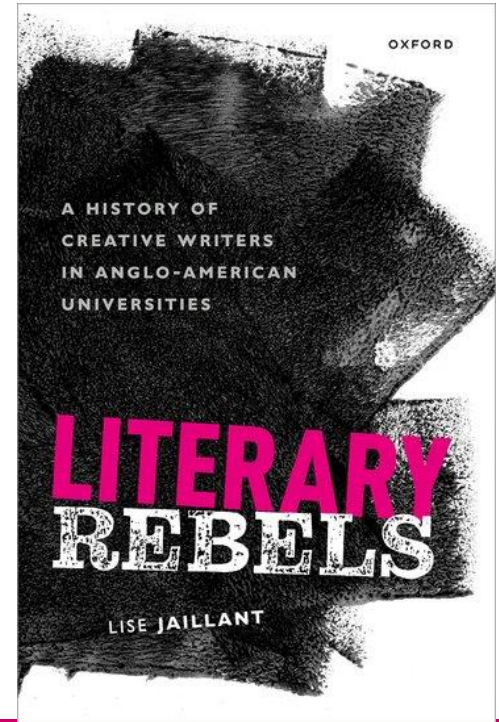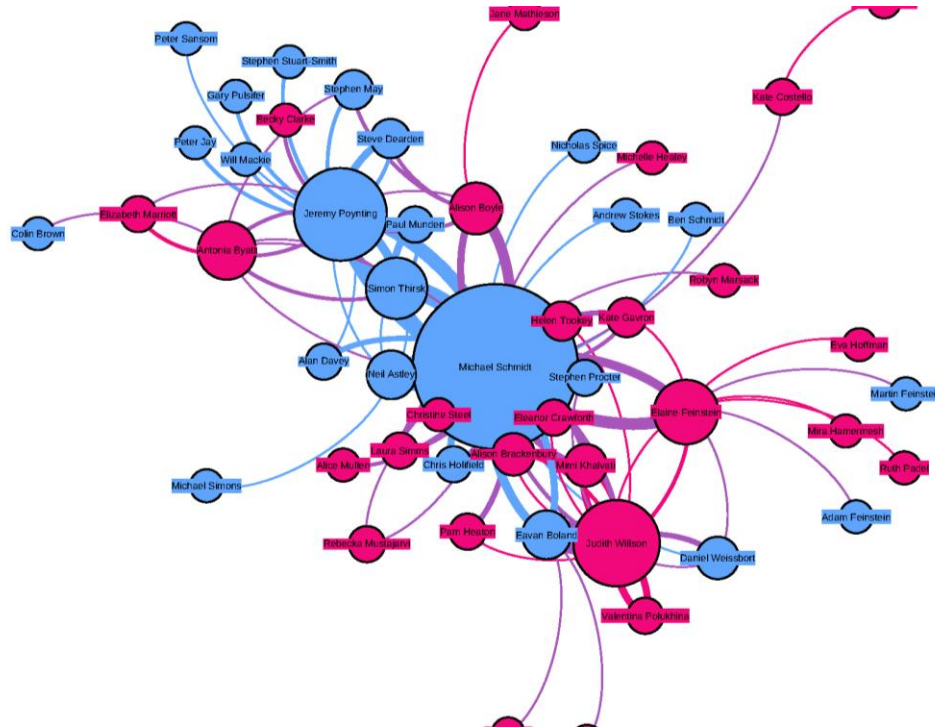*Sincerely,*
Seán Sweeney
Trustee

I am contacting you to request permission to consult the following documents:

University of Reading Library (Special Collections) > Jonathan Cape Archives > Travellers' Library:

1) **Production ledger (H-L): gives the print run for *Dubliners* and *A Portrait of the Artist as a Young Man* by James Joyce.**

2) **Three letters from Cape to Joyce in General correspondence file 1924 mentioning *A Portrait of the Artist as a Young Man* and *Ulysses*.**

Random House Group Archive & Library in Rushden:

3) **Contracts between James Joyce and Jonathan Cape**

Please find below a summary of my research:

The Modern Library, a cheap series of reprints marketed as the "world's best books," was created in New York in 1917. In the 1920s and 1930s, the series published texts by James Joyce, Virginia Woolf and Gertrude Stein alongside detective fiction and novels by writers that we now consider "middlebrow." My central argument is that the diversity of the Modern Library exemplifies the flexibility of cultural categories in the interwar period – a flexibility that was lost in the 1940s and 1950s when critics separated "high" cultural forms from "lesser" works. In this dissertation, I highlight the ways in which Random House sold the Modern Library to the academic market, thus contributing to the mainstreaming of modernism. I also compare the Modern Library and English publisher's series to show that these commercial entreprises created new markets for modern literature. Finally, I examine the construction of modern literature as a category radically different from "middlebrow" literary works. My dissertation is composed of three sections, two of which are completed. My ability to write the remaining section (Part II) on English publisher's series depends crucially on the research I will do in the United Kingdom in 2012-13.

Thank you in advance.
Sincerely,

Lise Jaillant

Letter from the
James Joyce
Estate,
10 March 2012

Loughborough University

Doing research in born-digital collections

IAN MCEWAN
ARCHIVE

@ HARRY RANSOM
CENTER (AUSTIN, TEXAS)

Credit: Spencer Murphy (https://www.telegraph.co.uk/books/authors/ian-mcewan-interview-lessons/)

OXFORD

A HISTORY OF
CREATIVE WRITERS
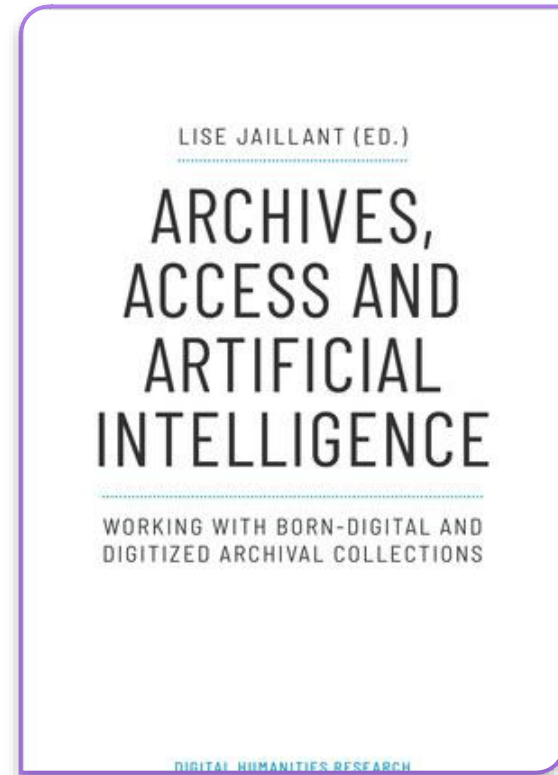IN ANGLO-AMERICAN
UNIVERSITIES

LITERARY
REBELS

LISE JAILLANT

Carcanet Press archive (MANCHESTER, UK)

Selection of CARCANET EMAILS

For more on Carcanet Press, see: "Design Thinking, UX and Born-digital Archives: Solving the Problem of Dark Archives Closed to Users"

In *Archives, Access and AI* (Transcript, 2022)

Available OPEN ACCESS

LISE JAILLANT (ED.)

ARCHIVES, ACCESS AND ARTIFICIAL INTELLIGENCE

WORKING WITH BORN-DIGITAL AND DIGITIZED ARCHIVAL COLLECTIONS

DIGITAL HUMANITIES RESEARCH

SUSAN SONTAG COLLECTION @ UCLA

c. 18,000 emails

- → Data protection issues

- → Includes emails in which she denies having a relationship with the photographer Annie Leibovitz

2014 article in *Los Angeles Review of Books*

"We also . . . know that Leibovitz was the witness to Sontag's will, and that Sontag's hard drives contain sensitive documents regarding Leibovitz's surrogate pregnancy. A simple keyword search for "Annie" on the UCLA laptop is illuminating, whether we're mongering for gossip or just trying to create an accurate portrait of Sontag's digital life."

"Sontag was committed to protecting her privacy during her lifetime, especially with regard to her sexuality, and thus it's surprising that she should choose to open her digital life to the public at all."

https://lareviewofbooks.org/article/excess-susan-sontags-born-digital-archive

Visit to UCLA (July 2022)

- Annie Leibovitz: [Sent messages: 8; Received messages: 10; Received from owner: 7]
- No indication to users that messages have been removed.
- Very limited selection, available only in the reading room = better than no access ?

Wendy Cope EMAIL COLLECTION @ British LIBRARY

25,000 individual emails

Callum McKean (Digital Lead Curator):

"Making sure that this material complies with data protection regulation in the UK before it is released is obviously a

considerable task."

https://blogs.bl.uk/english-and-drama/2020/04/born-digital-literary-archives-how-were-archiving-the-future.html

Section 2
IS AI A SOLUTION?

AI as a Solution?

- ✓ Sensitivity review

- 🔍 Identification of important records with historical value

- 🗄 Automatic creation of metadata

Externally-funded projects

Recent projects (since 2020) that I have led as UK PI:

AURA (Archives in the UK/ Republic of Ireland and AI: Bringing together Digital Humanists, Computer Scientists & stakeholders to unlock cultural assets) **– funded by AHRC and Irish Research Council.**

AEOLIAN (UK/ US: AI for Cultural Organisations) – **funded by AHRC and the US National Endowment for the Humanities.**

EyCon (Visual AI and Early Conflict Photography) **– funded by AHRC and LABEX in France**.

LUSTRE (Unlocking our Digital Past with AI) – funded by **AHRC Follow On Grant**, with the Cabinet Office as main partner

AEOLIAN (UK/ US: AI for cultural organizations)

www.aeolian-network.net

→ *UCL Press edited collection forthcoming in 2024*

→ *Special issue in Journal of Computing and Cultural Heritage*

Eycon (UK/ French project)

Unlocking digital photographs using AI

https://eycon.hypotheses.org/

→ Special issue in Digital Humanities Quarterly, forthcoming

Project in partnership with the Cabinet Office

➔*Focus on government archives*

https://lustre-network.net/

➔ *Special issue in AI & Society (forthcoming)*

Risks associated with AI

Addressing the digital heap @ Cabinet Office

"In one of our 'heaps' there were over **11 million files all of which needed to be reviewed** before they could either be retained as a historic record or destroyed. We estimated that **it would take one person 59 years to read them all**, and if we attempted to employ 59 people to complete the task in one year, the paybill would exceed £2million."

(David Canning, Cabinet Office)

https://codigital.blog.gov.uk/2023/01/16/conquering-the-digital-heap/

Conquering the 'digital heap'

In March 2022, across all its systems, the Cabinet Office held …

159 TB

If this information was printed out it would fill **103.3 Billion** pages of text, forming a pile 10,300 kilometers high
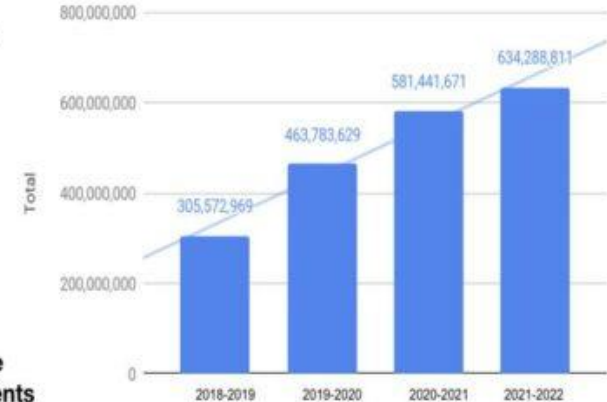
Of which .

75% is email

25% are documents

Total volume of digital objects (all information stores) 2018-19 to 2021-22

- 2018-2019: 305,572,969
- 2019-2020: 463,783,629
- 2020-2021: 581,441,671
- 2021-2022: 634,288,811

AI and the risk of censorship

- Machines making decisions on what to keep and what to delete
- Risk of errors: deleting important records without human oversight
- Need to keep humans in the loop

Inappropriate access to sensitive content:

Examples of medical images and colonial archives

# "Visually similar images" tool @Wellcome Collection

When you type in the word "skeleton" the search engine doesn't look at the images at all. It looks at the descriptions of those images – the captions, the titles, the whatever – and if a thing matches those titles, then you get a result. But the results are not always great, particularly in a collection of data that is as messy as ours. . . The example at the front of my mind is that there was a very kind of tight cluster of words that related to food and in the visual features that we extracted from images, **it was hard for the computer vision algorithm to differentiate between pictures of meat and pictures of gory surgery. An algorithm doesn't necessarily know how to differentiate that context.**

Anon., interview, 13 January 2023, via MS Teams.

Wellcome collection

https://wellcomecollection.org/works/wsqmrqfj
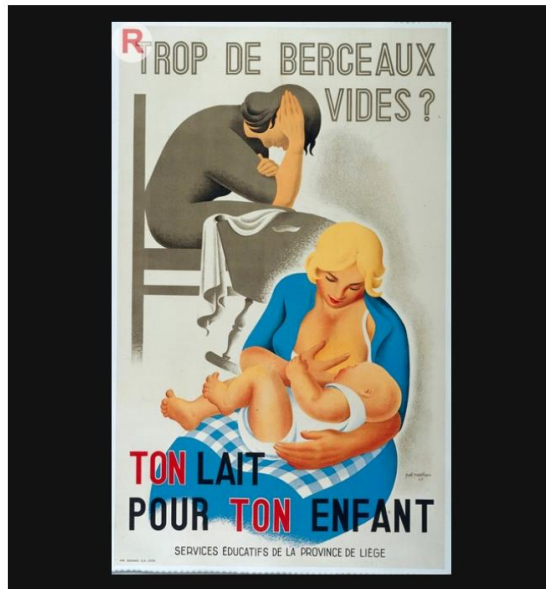
USE IMAGE NO. L0032741

Pol Mathieu

Licence: <u>In copyright</u>

⊙ **View image**

Visually similar images

We use machine learning to find images in our collection with similar shapes and features. <u>Let us know</u> if something doesn't look right.

Sensitivity and Access: Unlocking Colonial Visual Archives with AI

KEY CHALLENGES:

Many colonial archival photographs have been digitised without sufficient metadata

When metadata exists, it often uses problematic language

Sensitive materials from these collections often remain undigitised and hidden due in large part to ethical concerns

EyCon - Early Conflict Photography Database (1890-1918)

Avertissements

Ce matériel numérisé contient des informations et/ou des images qui peuvent :

-être perturbantes ou difficiles

-être explicites ou choquantes

-inclure des objets, des pratiques et des personnes décontextualisés d'une manière offensante pour la culture d'origine.

Take away points (DHQ article, forthcoming)

- Hiding sensitive images is not a viable option.

- AI can be used to make images – including sensitive images – more discoverable and accessible.

- However, **access cannot be separated from responsibility**. It is not responsible to release sensitive images without any oversight, or to push sensitive content to users who have viewed images considered similar.

- More national regulation and professional guidelines are needed to prevent children and vulnerable users to encounter sexually-explicit, violent and otherwise problematic images.

Section 3
CONCLUDING REMARKS

PROVIDING ACCESS RESPONSIBLY

- "archives are not Wikileaks, and we're not in the WikiLeaks business. … It's not responsible to data subjects; it's not responsible to other people's intellectual property rights; it's not lawful. So, we then need to build the techniques to provide access responsibly."

- *Interview with John Sheridan (The National Archives UK), 29 June 2021*

TRUSTING THE USER

- "I would like to see more focus on sharing that duty of care between the archives and the users. I think users need to take a far more active and conscious role in maintaining . . . the integrity of the archival discovery process. And I think that can be done by **having well developed and well-respected codes of conduct that the user very consciously agrees with beforehand**, which means that even if they see sensitive information, that sensitive information goes no further than that individual researcher."

- *Interview with Adam Nix (academic), 22 June 2021*

1. Consider giving access only to trusted users, but make this selective access transparent

2. Explore options to give access remotely

3. Encourage two-way discussions between archivists and users

Practical suggestions

Digital Humanities

- **Digital Humanities – which is in essence an interdisciplinary field – has an important role to play in bridging the gap between record creators, archivists, researchers, and other users.**

Further readings (see www.lisejaillant.com)

"How Can We Make Born-Digital and Digitised Archives More Accessible? Identifying Obstacles and Solutions." *Archival Science* 22 (2022): 417-36. *OPEN ACCESS*

"Unlocking Digital Archives: Cross-disciplinary Perspectives on AI and Born-digital Data" co-authored with Annalina Caputo. *AI & Society* 37 (2022): 823-35. *OPEN ACCESS*