

---

# AI and Archives from a Researcher's Viewpoint

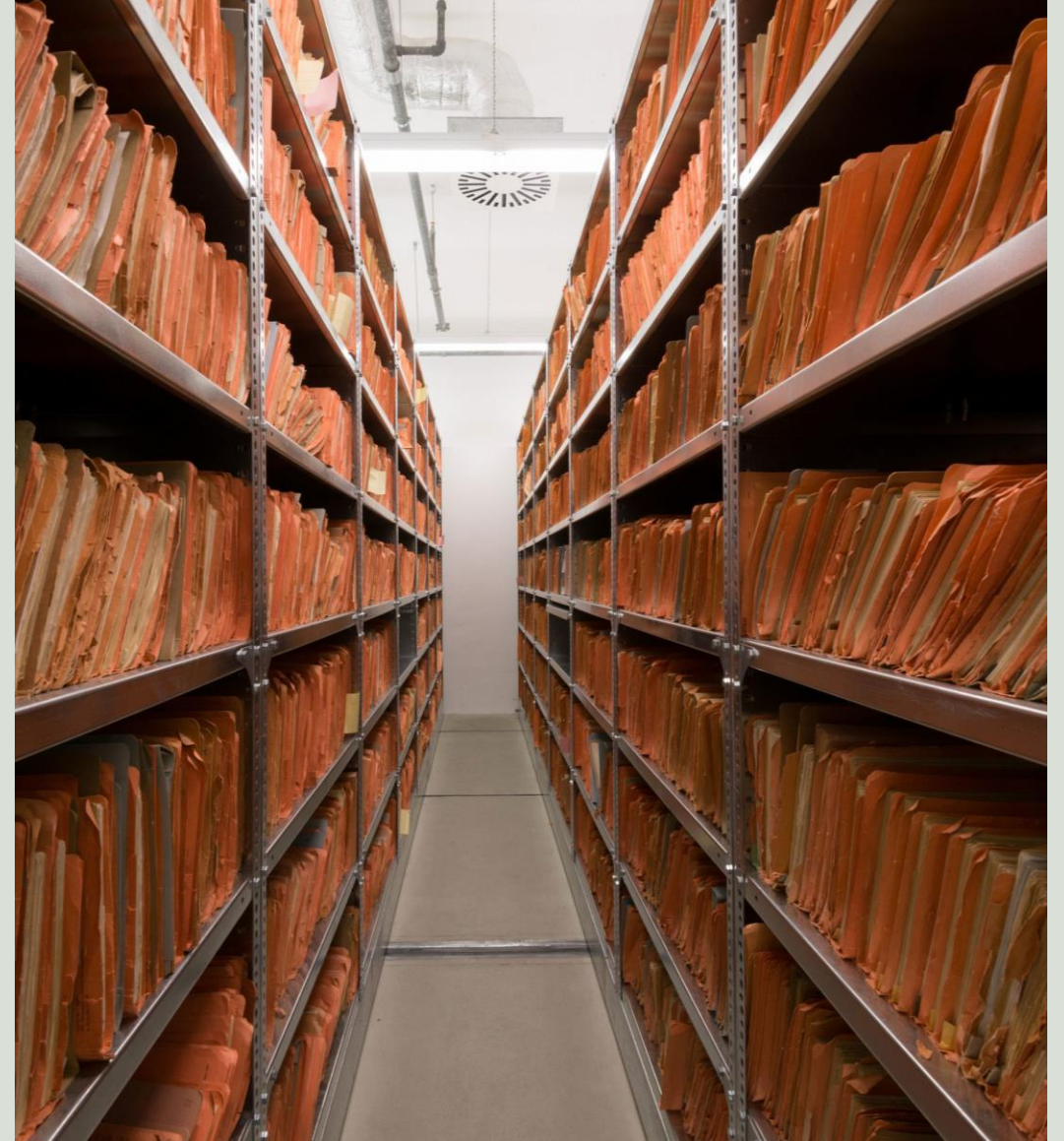
*Lise Jaillant*

*Loughborough University*

---

---

# From traditional archives to born-digital archives



# Experience

Recent projects I have led as UK PI:

- (1) **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.**
- (2) **AEOLIAN** (UK/ US: AI for Cultural Organisations) – **funded by AHRC and the US National Endowment for the Humanities.**
- (3) **EyCon** (Visual AI and Early Conflict Photography) – **funded by AHRC and LABEX in France.**
- (4) **LUSTRE** (Unlocking our Digital Past with AI) – funded by **AHRC Follow On Grant**, with the Cabinet Office, The National Archives and other partners

---

## THE PROBLEM

Policy makers produce digital records on a daily basis: emails, PDFs, Word documents, A/V files, etc.

A selection of records is then preserved in archival repositories.

However, getting access to these archival materials is extremely complicated for many reasons – including data protection, sensitivity, national security, and copyright.

---

## THE SOLUTION?

**Artificial Intelligence can be applied to archives to make them more accessible, but it is still at an experimental stage.**

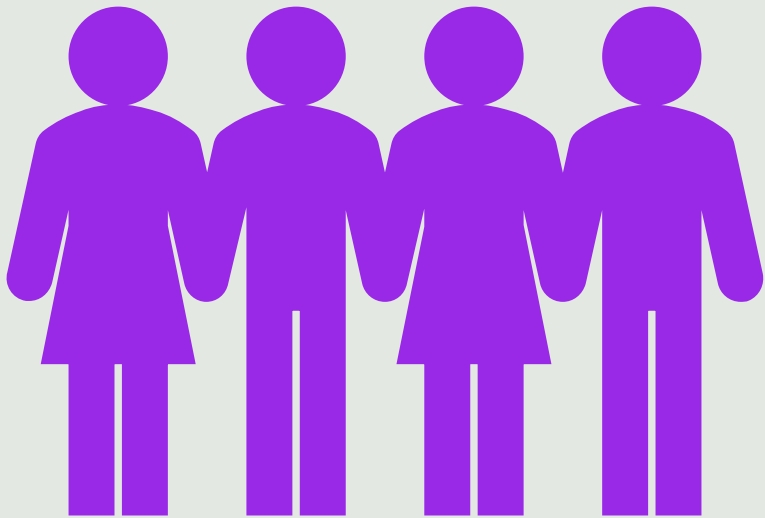
**Issues with skills gaps.**


**BUT also issues of mistrust and miscommunication.**

---

**mistrust between  
stakeholders**

**mistrust of technology**





"What do you need to make a new AI system? You need algorithms, you need people, you need computer, you need data. And for this AI system to also be useful in the world, you need funding and you need product market fit. And you need to have sufficient levels of trust and performance."

(Interview with Shahar Edgerton, Centre for the Study of Existential Risk [Cambridge], 3 Nov. 2022)



---


**What are the potential applications of AI to make born-digital archives more accessible and usable?**

**And what are the barriers to implementing them?**





---



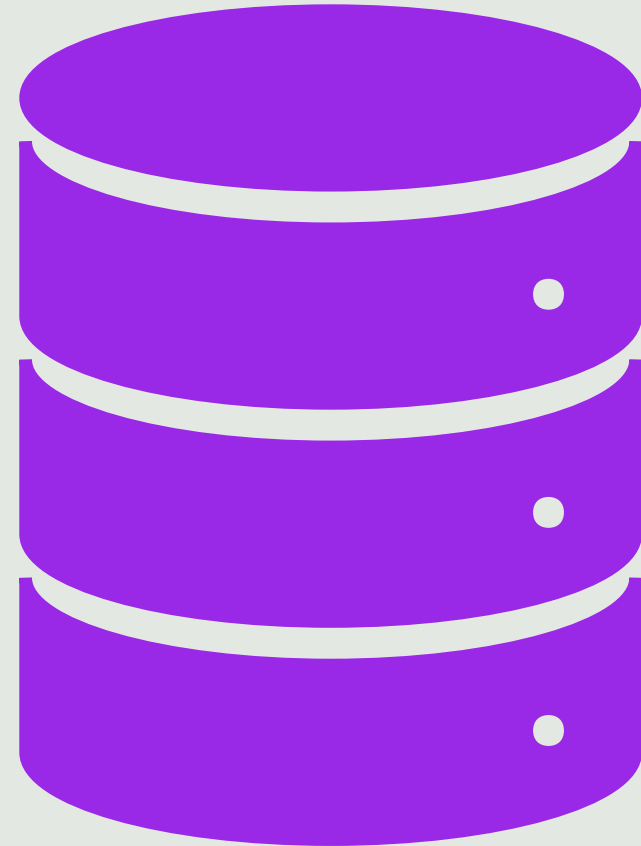
Our argument  
stakeholders share  
similar professional  
ethics

→ *Surfacing these similarities can lead to deeper  
collaborations between humans*

---

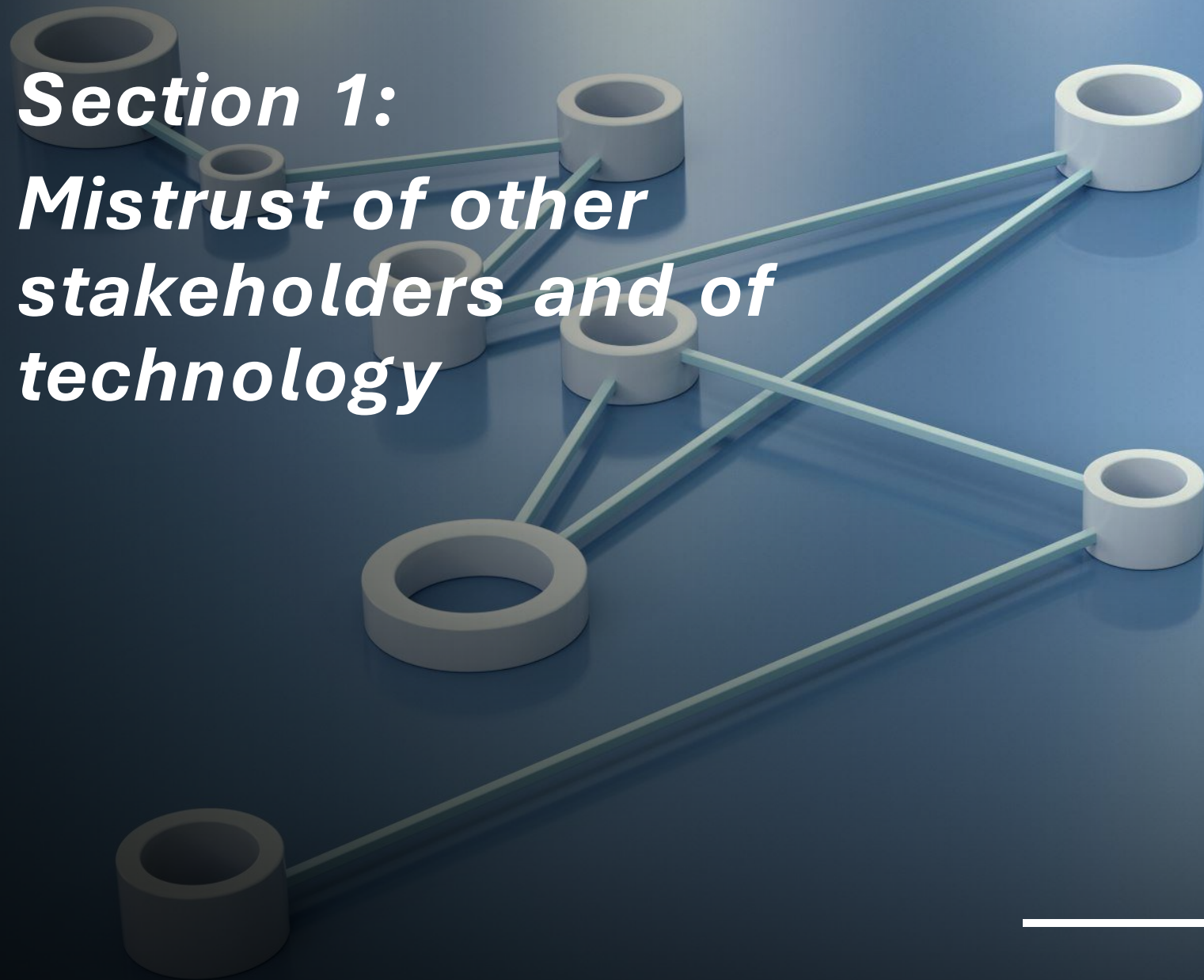
# TRUST AND COLLABORATION

**archival circle (from  
record creators to  
archivists, and from  
archivists to users)**



---

***Section 1:***  
***Mistrust of other  
stakeholders and of  
technology***





## A DIGITAL DARK AGE?

“I always feel very sorry for my archivist friends every time historians turn round and say, well, you must collect everything, and why aren’t you trying, you know, we need all this stuff and it’s going to be a digital dark age. And they’re like, **it’s not going to be a digital dark age, we always do selection**, you know, it’s just that selection issues are kind of way more complicated now in the age of the born digital.”

*Interview with James Baker (academic, formerly GLAM sector professional), 22 June 2021*



## SCALE OF BORN-DIGITAL ARCHIVES

“The National Archives in the US today has 500 million emails from the Regan administration to the Obama administration. The Trump emails haven’t been counted yet or haven’t been fully processed. But it has 500 million distinct emails, and then more than a billion pages because those emails have attachments.”

*Interview with Jason Baron (academic/ legal professional), 22 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*



## 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.”

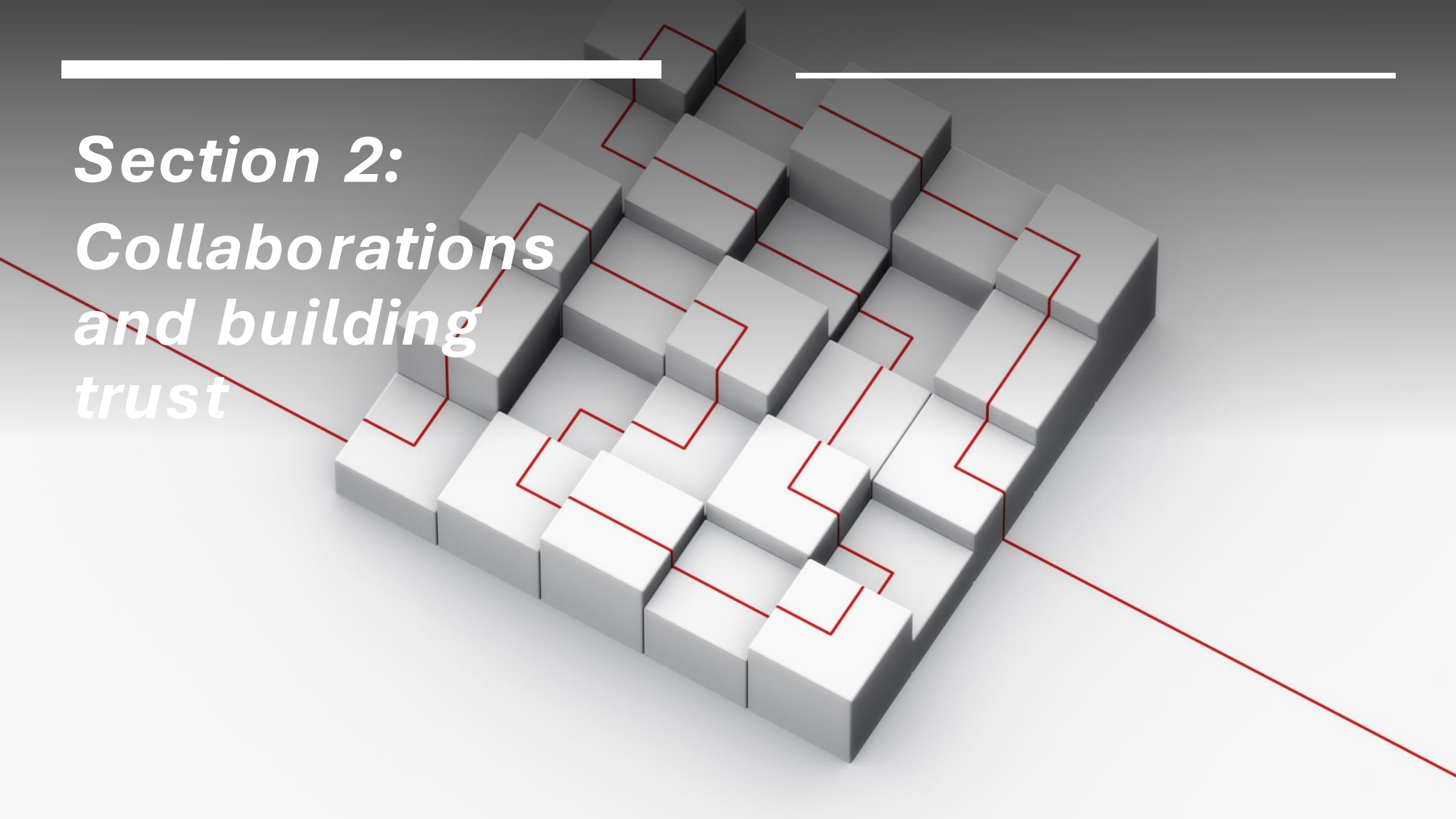
*Interview with John Sheridan (The National Archives UK), 29 June 2021*



---

---

## *Section 2: Collaborations and building trust*

A 3D isometric illustration of a grid of cubes. The cubes are arranged in a roughly rectangular pattern, with some cubes missing or offset to create a stepped effect. The cubes are colored in a gradient from light gray to white. Red lines are drawn across the top surfaces of the cubes, forming a network of interconnected paths. A single red line also extends from the bottom left corner of the grid towards the bottom right corner of the frame.



## KNOWLEDGE TRANSFER

“It’s all about the knowledge transfer into our sector from outside. No one is building anything specifically for us. It is all about things that are being built in the world, and then someone’s going, “Oh, I could apply this to archives.” And I think all we can really do is equip ourselves with learning to understand what that kind of technology transfer process looks like and understand what the implications are.”

“it’s quite hard for us to know how to navigate this world where, whilst we have collections, we’re effectively, certainly in terms of budgets, and then in terms of capability, we’re fleas around elephants. But we’re not on our own. These issues are pervasive, and again, I think it’s about not just talking to each other, but talking to other people who share concerns, is going to be very important, too.”


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



## ENGAGING WITH ACADEMICS AND OTHER USERS

“we have had a few attempts at trying to run events for researchers to find out what they might be expecting, or what they might want. And . . . the majority of our existing core research audience is not yet working with born-digital material.”

*Interview with GLAM sector professional [name withheld], 27 May 2021*



“I thought [the researchers] would be much more self-confident, but they all seemed to doubt their essential skill sets as historians in using this kind of material and, for similar reasons, hoping to be retired too before they had to do too much of this, which really surprised me.”

*Interview with Andrew Riley (archivist), 7 July 2021*



## FACILITATING ACCESS

“When you’re talking about the public sector, the work we do is funded by public budgets or scrutinised by Parliament and by journalists and by everybody else. **Most of our work should be public by default.** . . . I think if more of our work was public, I think more people would find it, kind of, boring. But also, it wouldn’t be [in case of data leaks] “Ooh, secret trove of Cabinet Office emails released by aide.”

*Interview with government professional, 1 July 2021*



there can be no betrayal without trust.



It is precisely the lack of trust in other stakeholders, and in new technologies, that was highlighted in our interviews.



Mistrust of technology makes it difficult to implement AI tools.



The concerns highlighted by interviewees – in terms of accountability, control, and bias – make it difficult to accelerate the development of AI applied to archives.



---

What is needed:

---

more cross-sector and cross-disciplinary dialogue to surface the shared professional ethics between stakeholders.

---

more research across disciplinary boundaries.



---

# 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.**

---

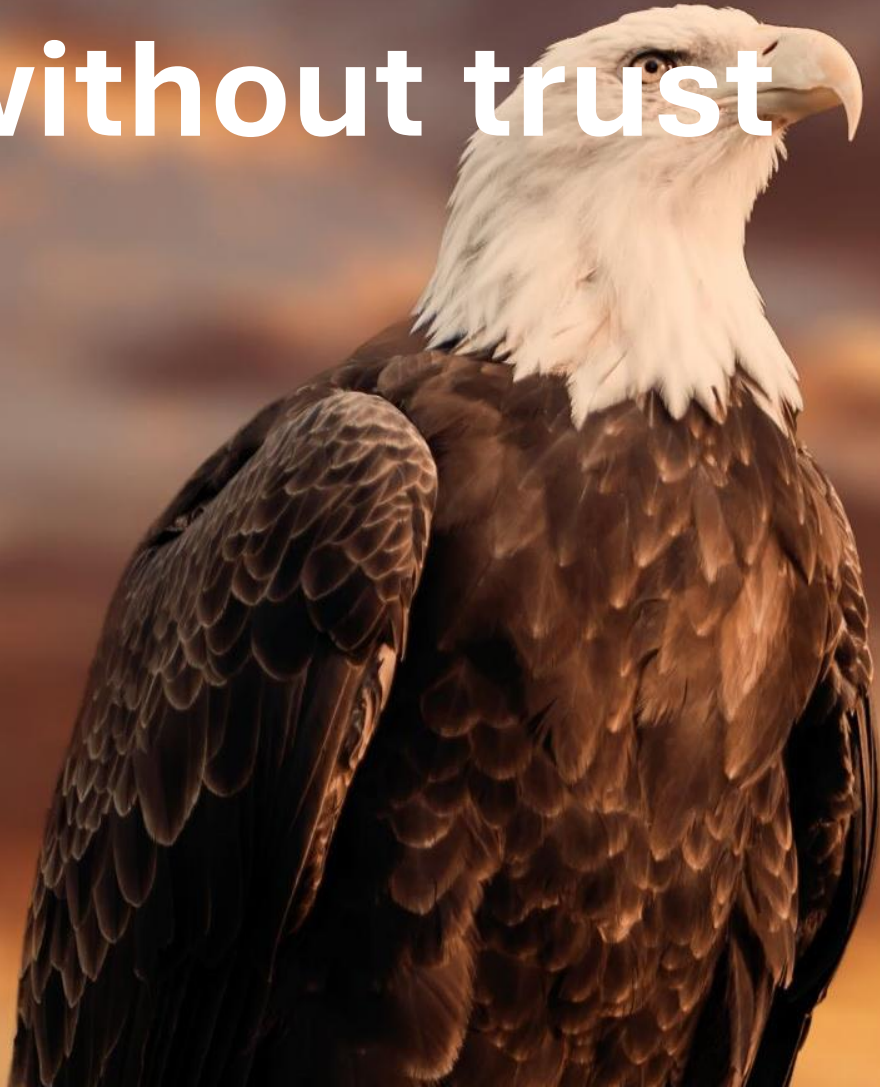
**“For there to be betrayal, there would  
have to have been trust first.”**

**Suzanne Collins**

***The Hunger Games***



**there can be no betrayal  
without trust**





Thank you!

[l.jaillant@lboro.ac.uk](mailto:l.jaillant@lboro.ac.uk)