

# **Finding light in dark archives: Using AI to connect context and content in email**

**Dr Adam Nix**

**De Montfort University**

***Co-I, Contextualising Email Archives***

**AURA: Artificial Intelligence and Archives: What comes next?**

# A user perspective

- Born-digital sources are vital for future historical research
- Many born-digital collections remain inaccessible (dark)
- Methodology largely assumes the use of original, pre-digital sources
- Ethical implications of using digital sources are unclear

# Beyond Preservation

- Born-digital sources are shaking up traditional archival processes (Prom et al., 2019)
- Managing privacy at scale remains a key barrier to born-digital access (Milligan, 2019)
- Answering the problem of access requires collaboration ‘between both sides of the reading room’ (Jaillant, 2019)
- Users value optionality, expect some curation, but are still inexperienced (Welcome, 2017)



How will users *actually* engage with born digital material once access issues have been navigated?

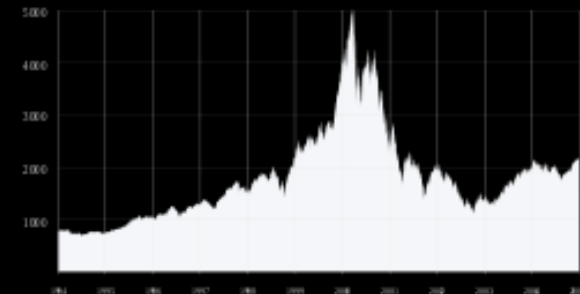
# Contextualizing Email Archives

Explores the gap between current efforts to preserve emails, and the means by which researchers might actually read and engage with them.



- Uses the emails of a failed dotcom company
  - Used under the pseudonym “Avocado-IT”
  - Preserved and made available for research via the LDC

Our project explores new ways to make an email archive available to search and study while maintaining the relational and network properties



# Email specific issues

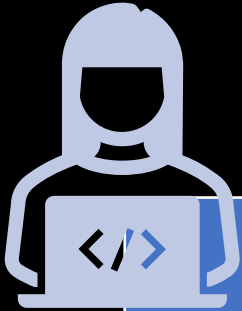
- The networked nature of organizational email make it difficult to search effectively
  - Email is a hybrid artifact: email IS and email ARE
  - Not just information as content, but also as context
- Non-historians often engage with just one aspect of this networked resource:
  - Frequency and networks
  - Timing and sequencing
  - Language
  - Content

*For organizational email to become a useful historical source, both individual and network aspects of email need to be maintained.*

## We also assume a need to...

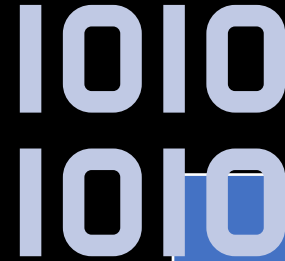
- Accommodate increasingly diverse research questions
- Allow users to work iteratively through a collection
- Work with the tacit (sometimes messy) nature of historical research
- Provide for different levels of experience
- Offer relatively complete access to a whole organisational corpus

# Approaches to contextualization



- **User-led**

- Reliant on search interface
- Selection based on email's structural characteristics
- Searching with keywords
- Requires clear and parsed queries

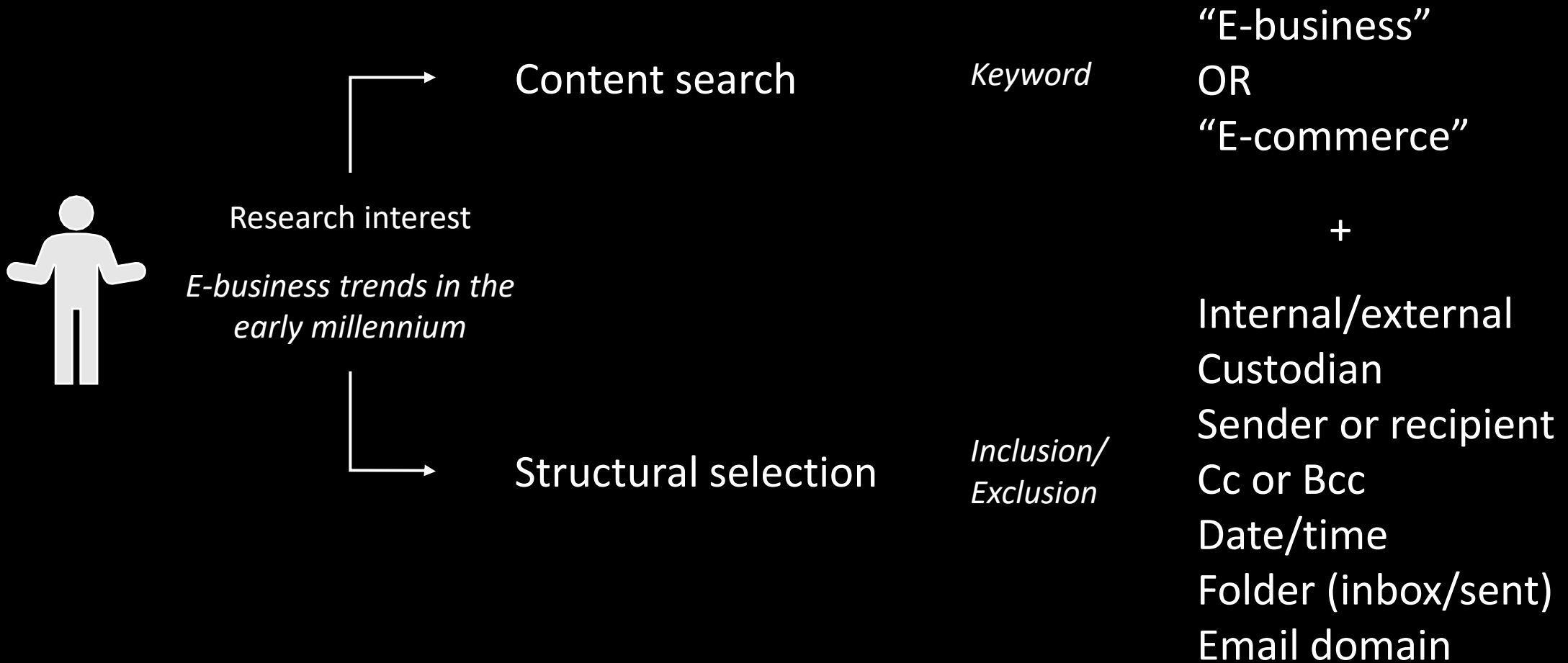


- **Machine-led**

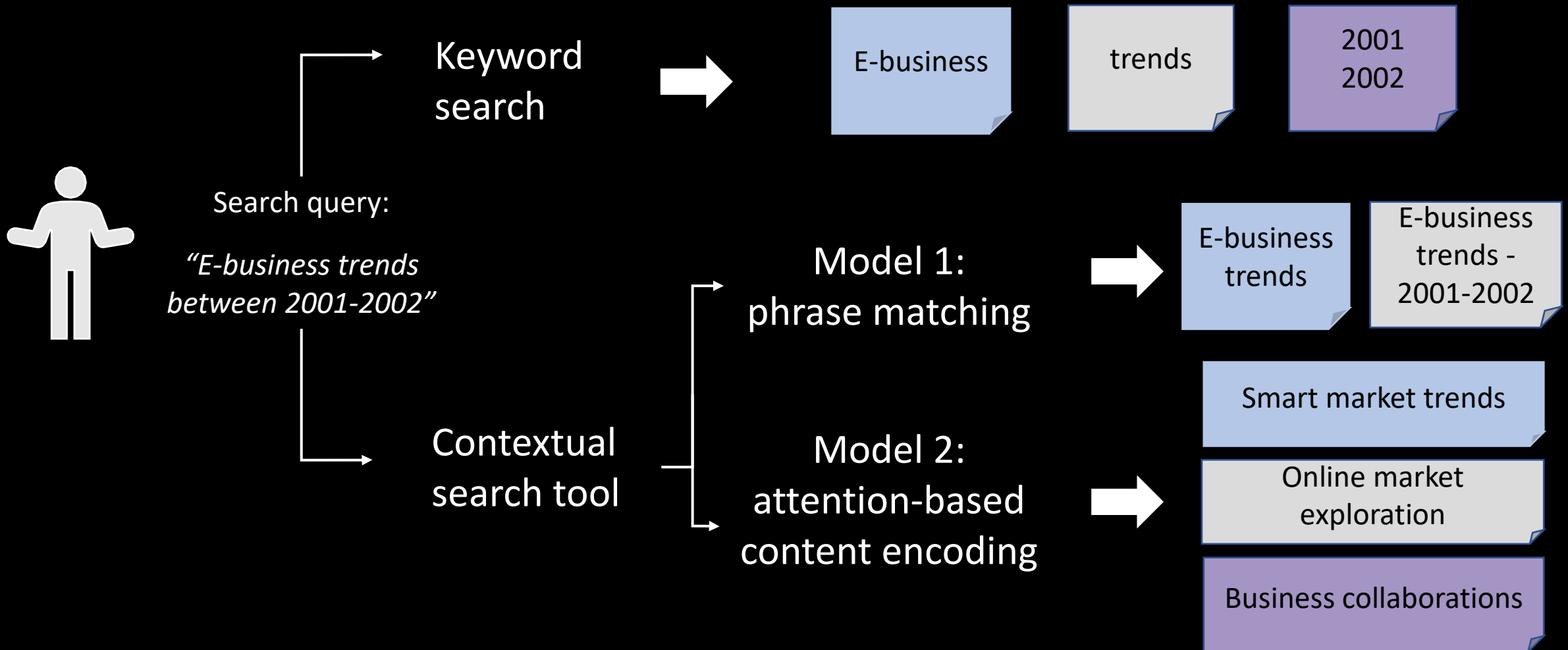
- Reliant on algorithmic processes
- Allows for 'simple' search
- New contextual discovery potential
- Raises questions around source criticism

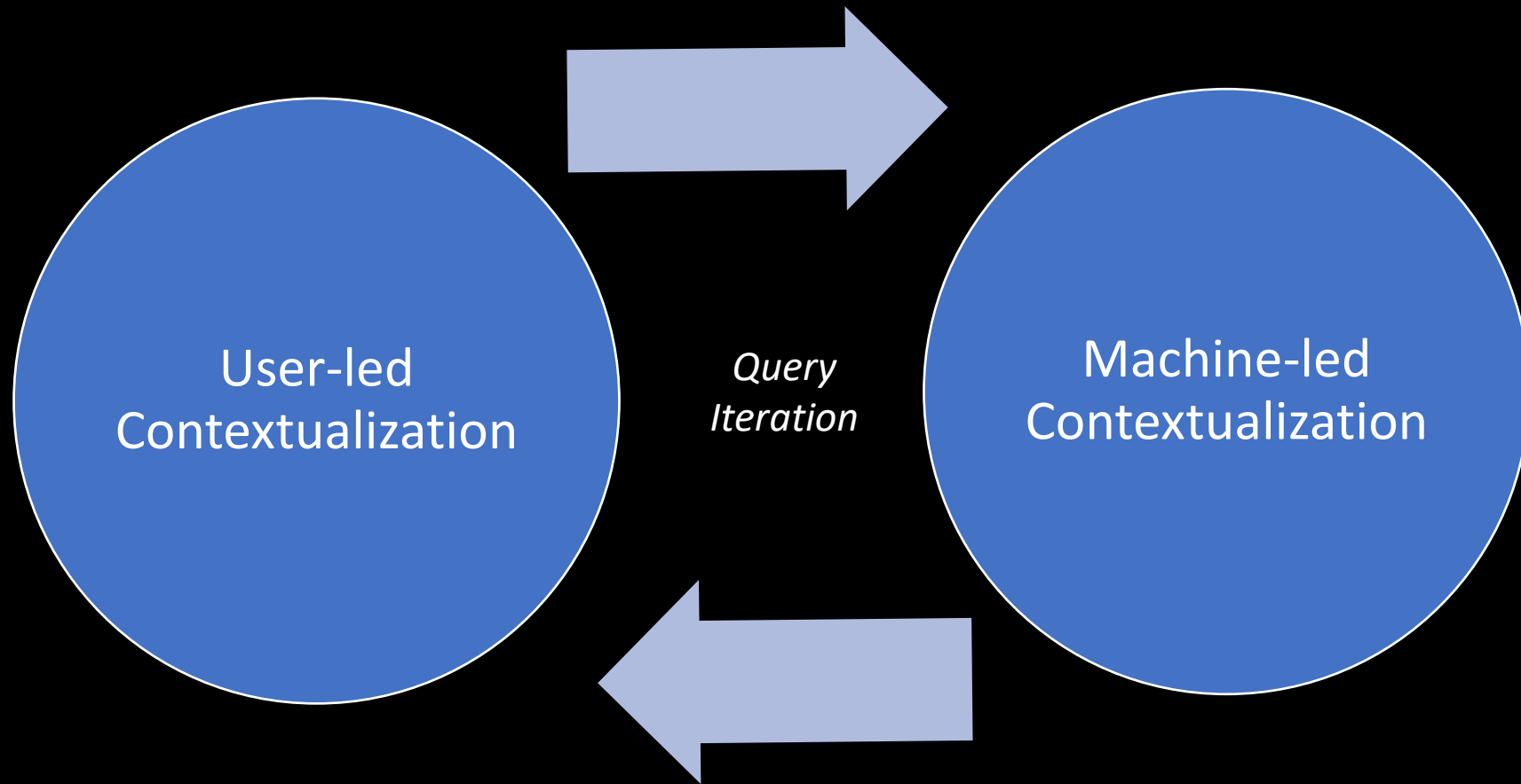


# User-led Contextualization



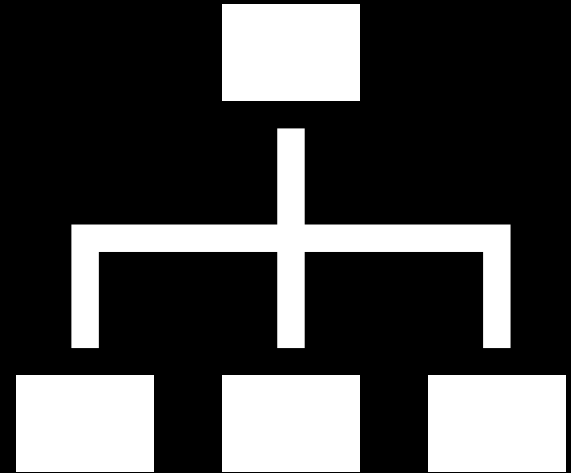
# Machine-led Contextualization





# Bringing context in

- External sources of context:
  - Organizational chart (structure)
  - Known relationships (CEO and PA)
  - Market or geographic events
  - Organizational records/correspondence



# What comes next?

- Finding out how historians will actually use born-digital:
  - What are the historical-use implications of unfamiliar archival technologies (including artificial intelligence)?
  - How will new practices and approaches be integrated into historical methods?
  - Could more 'beta' access to collections facilitate user experience and interface development?

*Thank you!*

## Historicizing the dot.com bubble & Contextualizing email archives

- Prof Stephanie Decker (PI) *University Of Bristol*
- Dr Adam Nix (Co-I) *De Montfort University*
- Dr Santhilata Kuppili Venkata (Co-I) *The National Archives*
- Prof David Kirsch (Co-I) *University Of Maryland*
- Dr Eric Rau (Heritage Partner) *Hagley Museum And Library*



Arts and  
Humanities  
Research Council