Privacy for Data Explorers in the 21st Century

Yunhyong Kim & Zoe Bartliff

Lecturers in information studies
University of Glasgow

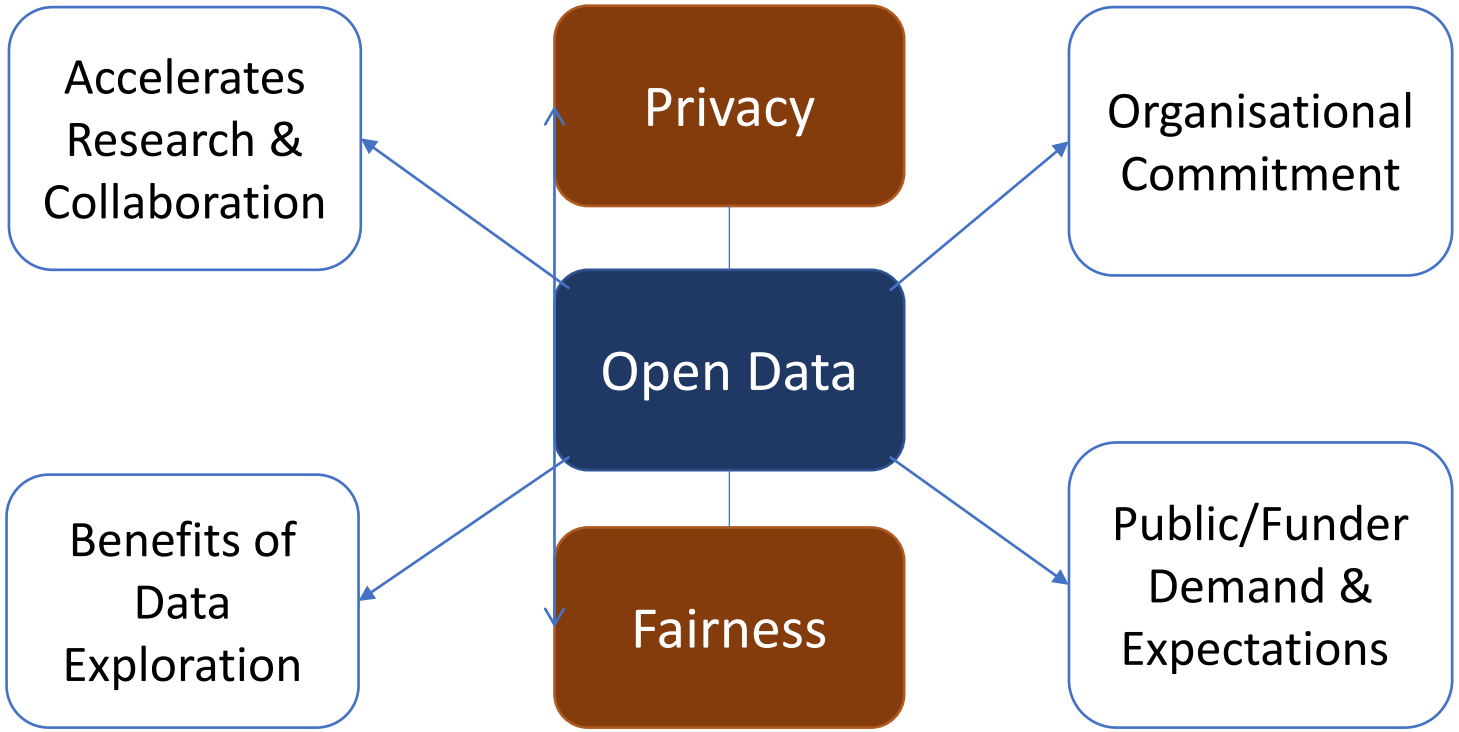
Yunhyong.Kim@glasgow.ac.uk

Zoe.Bartliff@glasgow.ac.uk

Outline

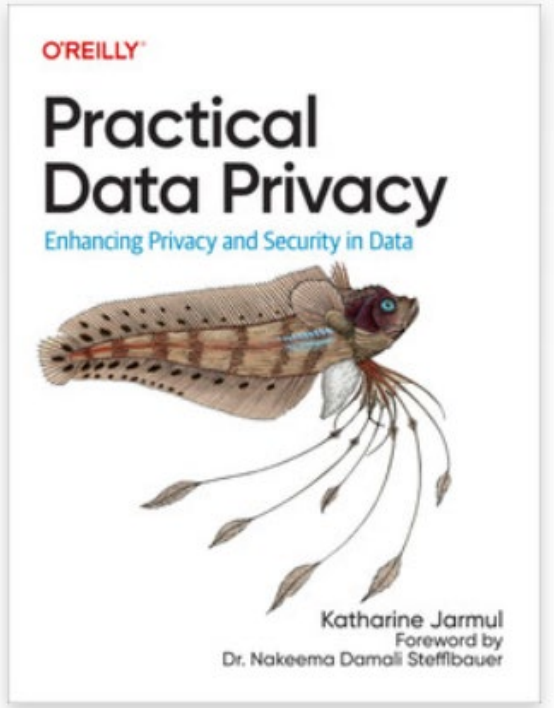
- Overview of Privacy
  - AI and Privacy
  - Archives and Privacy
- Case study: email archives
- The future

Open Data: a Common Goal



Practical Data Privacy (Katharine Jamul 2023)

- Understanding regulations like GDPR (EU/UK) – what does it mean for my data workflows and data science use cases?
- Critically thinking about strategies for privacy-aware data processing:
  - Anonymisation, aggregation, differential privacy
  - Federated learning and analysis
  - Homomorphic encryption
- How do I compare and choose the best privacy-preserving technologies and methods? Are there open-source libraries that can help?
- Can I ensure that my data science projects are secure by default and private by design?
- Working with governance and information security teams to plan your strategies in the project.



https://www.oreilly.com/library/view/practical-data-privacy/9781098129453/

Data Breach

British Library: Employee data leaked in cyber attack

21 November 2023

https://www.bbc.co.uk/news/entertainment-arts-67484639

British Library hack: Customer data offered for sale on dark web https://www.bbc.co.uk/news/entertainment-arts-67544504

Capita cyber-attack: USS pension fund members' details may have been stolen

Universities Superannuation Scheme says it can not be certain information about 470,000 members is safe

Rob Davies

@ByRobDavies

Fri 12 May 2023 16.29 BST

The Universities Superannuation Scheme (USS) invests almost £90bn on behalf of academics. Photograph: Anna Stowe/Alamy

Almost half a million members of the giant USS lecturers' pension fund may have had their personal details stolen during the recent cyber-attack on the outsourcing firm Capita.

https://www.theguardian.com/business/2023/may/12/capita-cyber-attack-uss-pension-fund-members-details-may-have-been-stolen

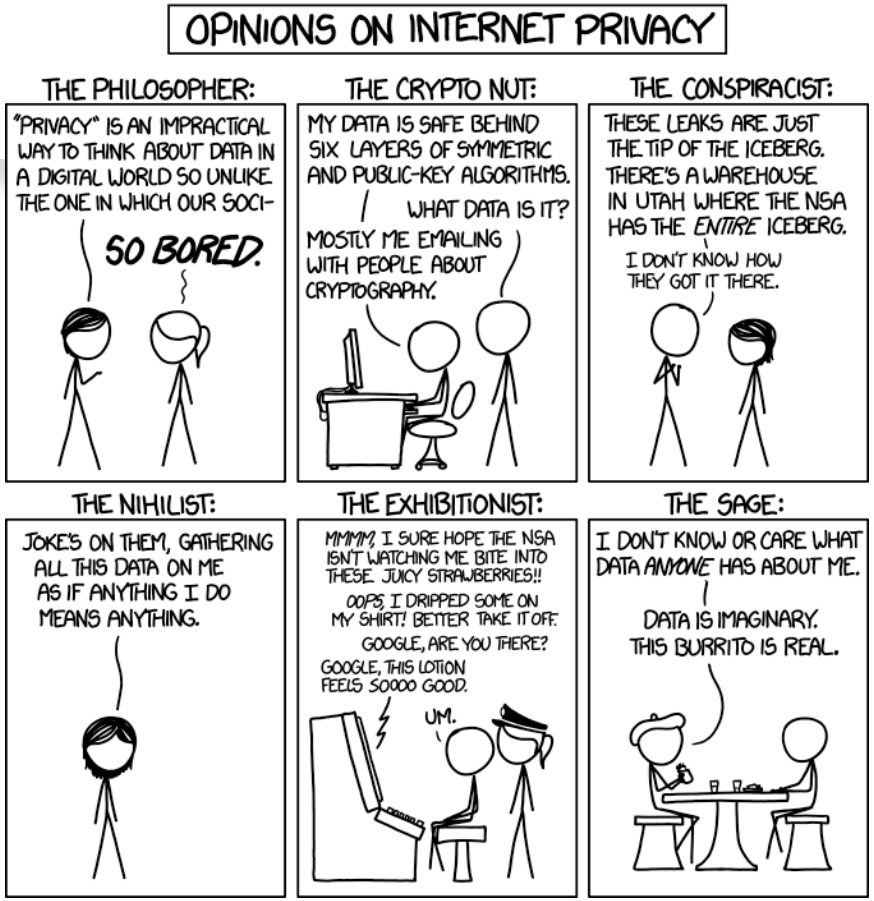
"members' titles, initials, names, dates of birth, National Insurance numbers and pension fund membership numbers"

Privacy in a Nutshell

It is **related to but not identical to data protection**. So it is not only about those protected characteristics mentioned by, for example, the General Data Protect Regulations (GDPR). It is about the broader ethics and human rights to be able to control their own data.

It is **related to but not identical to information security**. So it is not only about controlling who can access different types of data. It is about information governance that guides legitimate and ethical use, distribution, and storage of data throughout the data lifecycle.

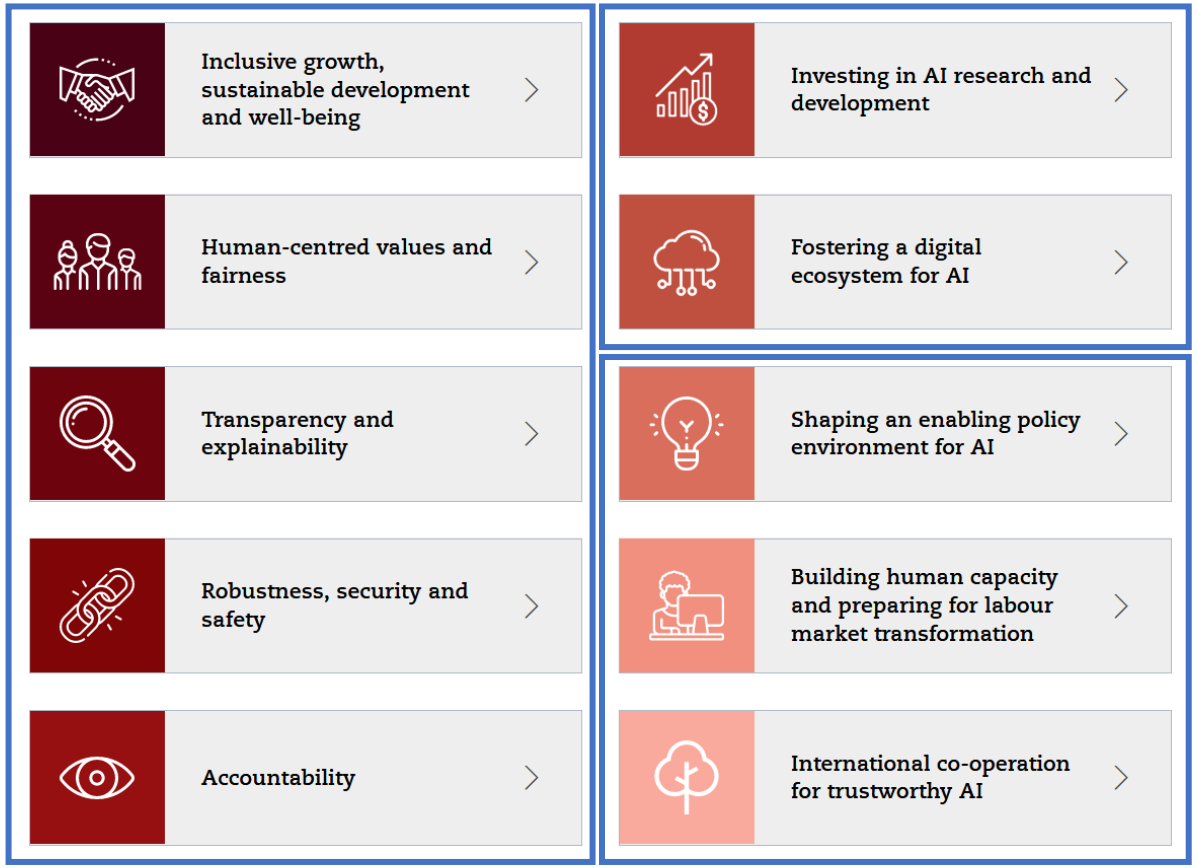
It is **related to but not identical to considering your intentions regarding information about an individual**. It is about considering the impact of the way you use data to the individual, those related to them, and social/cultural consequences.



AI and Privacy

AI Spotlight on Privacy and Fairness (1)

- Global Partnership on AI (GPAI)
  - International initiative to promote responsible AI use that respects human rights and democratic values
  - Conceived by Canada and France with 13 other founding members: Australia, the European Union, Germany, India, Italy, Japan, Korea, Mexico, New Zealand, Singapore, Slovenia, the United Kingdom and the United States.
  - Built on OECD AI Policy Observatory Principles
- Google AI Principles
- AI and Machine Learning for Coders (Laurence Moroney) - 2020
  - Chapter 20. AI Ethics, Fairness and Privacy
- Building an Anonymization Pipeline (Luk Arbuckle, Khaled El Emam) - 2020
- Building Machine learning pipelines (Hannes Hapke; Catherine Nelson)
  - Chapter 14. Data Privacy for Machine Learning
- The AI Ladder (Rob Thomas, Paul Zikopoulos) – 2020
  - Chapter 3. How to Overcome AI failures and Challenges



https://oecd.ai/en/ai-principles

AI Spotlight on Privacy and Fairness (2)

- Global Partnership on AI (GPAI)
  - International initiative to promote responsible AI use that respects human rights and democratic values
  - Conceived by Canada and France with 13 other founding members: Australia, the European Union, Germany, India, Italy, Japan, Korea, Mexico, New Zealand, Singapore, Slovenia, the United Kingdom and the United States.
  - Built on OECD AI Policy Observatory Principles
- Google AI Principles
- AI and Machine Learning for Coders (Laurence Moroney) - 2020
  - Chapter 20. AI Ethics, Fairness and Privacy
- Building an Anonymization Pipeline (Luk Arbuckle, Khaled El Emam) - 2020
- Building Machine learning pipelines (Hannes Hapke; Catherine Nelson)
  - Chapter 14. Data Privacy for Machine Learning
- The AI Ladder (Rob Thomas, Paul Zikopoulos) – 2020
  - Chapter 3. How to Overcome AI failures and Challenges

- Google AI principles:
  - Socially beneficial
  - Avoid creating or reinforcing unfair bias
  - Built and tested for safety
  - Be accountable to people
  - Incorporate privacy design principles
  - High standards of scientific excellence
  - Be made available for uses that accord with these principles
- Fairness
  - Is ML actually necessary
  - Design and implement metrics from day one
  - Build a minimum viable model and iterate
  - Infrastructure that supports rapid redeployment
- Explainability: e.g. What-If tool
- Privacy
  - Federated learning: data never leaves your device, model shared

AI Spotlight on Privacy and Fairness (3)

- Global Partnership on AI (GPAI)
  - International initiative to promote responsible AI use that respects human rights and democratic values
  - Conceived by Canada and France with 13 other founding members: Australia, the European Union, Germany, India, Italy, Japan, Korea, Mexico, New Zealand, Singapore, Slovenia, the United Kingdom and the United States.
  - Built on OECD AI Policy Observatory Principles
- Google AI Principles
- AI and Machine Learning for Coders (Laurence Moroney) - 2020
  - Chapter 20. AI Ethics, Fairness and Privacy
- Building an Anonymization Pipeline (Luk Arbuckle, Khaled El Emam) - 2020
- Building Machine learning pipelines (Hannes Hapke; Catherine Nelson)
  - Chapter 14. Data Privacy for Machine Learning
- The AI Ladder (Rob Thomas, Paul Zikopoulos) – 2020
  - Chapter 3. How to Overcome AI failures and Challenges

- Privacy and AI: could cause reduced accuracy and cost in time
- Privacy and fairness: privacy (e.g. deletion of personal information) could affect the possibility of gauging fairness, e.g., bias of data
- Privacy and open data: closed parts of the data could reduce the utility of the data
- Complexity of privacy: identifiability is not just about direct collection of personal data
- Data privacy: key questions
  - Who are you trying to keep the data private from
  - Which parts of the system can be private and which can be exposed to the world
  - Who are the trusted parties that can view the data
- Control of data passed to creators
- Encrypted machine learning
- Scrubbing the data for example using regex and named entity recognition
- Differential privacy: emulating distribution but not individual instances
- Federated learning: only sharing models – i.e. data never leaves home.

Failures and Opportunities

- Global Partnership on AI (GPAI)
  - International initiative to promote responsible AI use that respects human rights and democratic values
  - Conceived by Canada and France with 13 other founding members: Australia, the European Union, Germany, India, Italy, Japan, Korea, Mexico, New Zealand, Singapore, Slovenia, the United Kingdom and the United States.
  - Built on OECD AI Policy Observatory Principles
- Google AI Principles
- AI and Machine Learning for Coders (Laurence Moroney) - 2020
  - Chapter 20. AI Ethics, Fairness and Privacy
- Building an Anonymization Pipeline (Luk Arbuckle, Khaled El Emam) - 2020
- Building Machine learning pipelines (Hannes Hapke; Catherine Nelson)
  - Chapter 14. Data Privacy for Machine Learning
- The AI Ladder (Rob Thomas, Paul Zikopoulos) – 2020
  - Chapter 3. How to Overcome AI failures and Challenges

- Data challenges
  - Data silo
  - Lack of data
  - Disorganised data
  - Data quality
  - FAIR principle (findable, accessible, interoperable, reusable)
- Cultural Challenge
  - Potential employees with knowledge, skillset and experience are rare
  - Company culture and organisational silos
- Building trust
  - Fully traceable provenance
  - Lineage of the model and training data
  - Inputs and outputs of AI recommendation
  - Explainability

Situation in Archives

A gap in the debate about access, AI and privacy

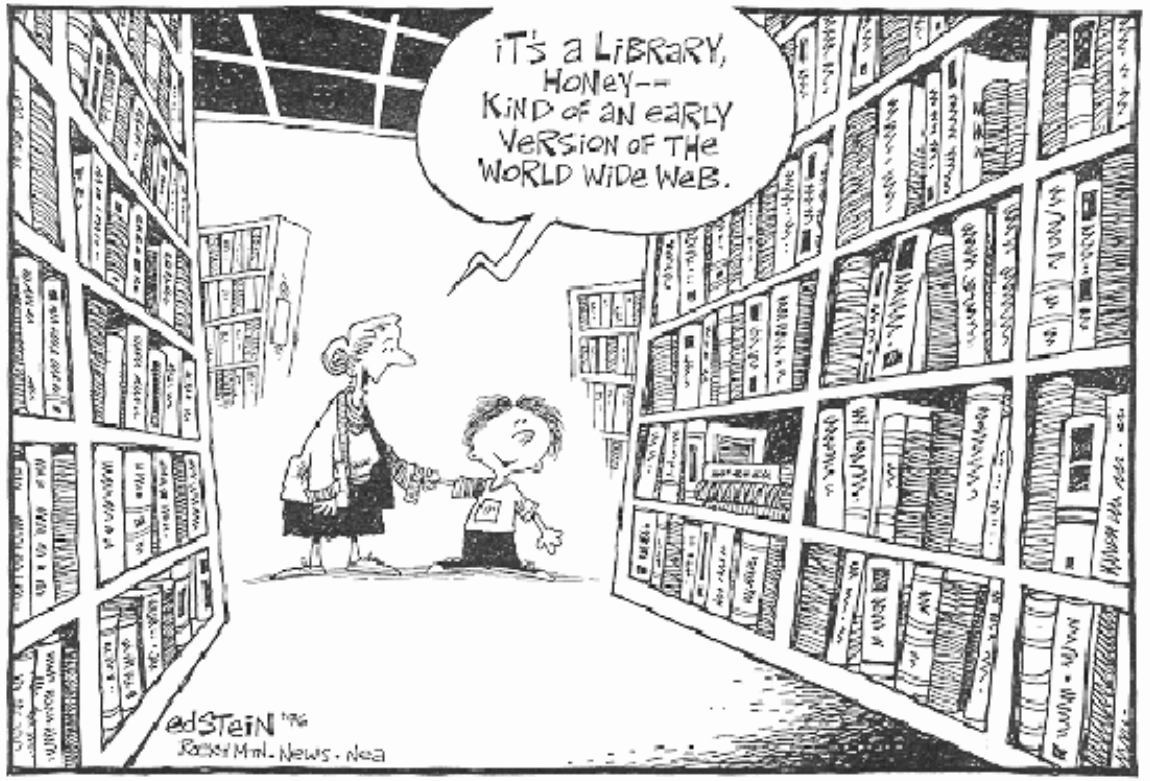
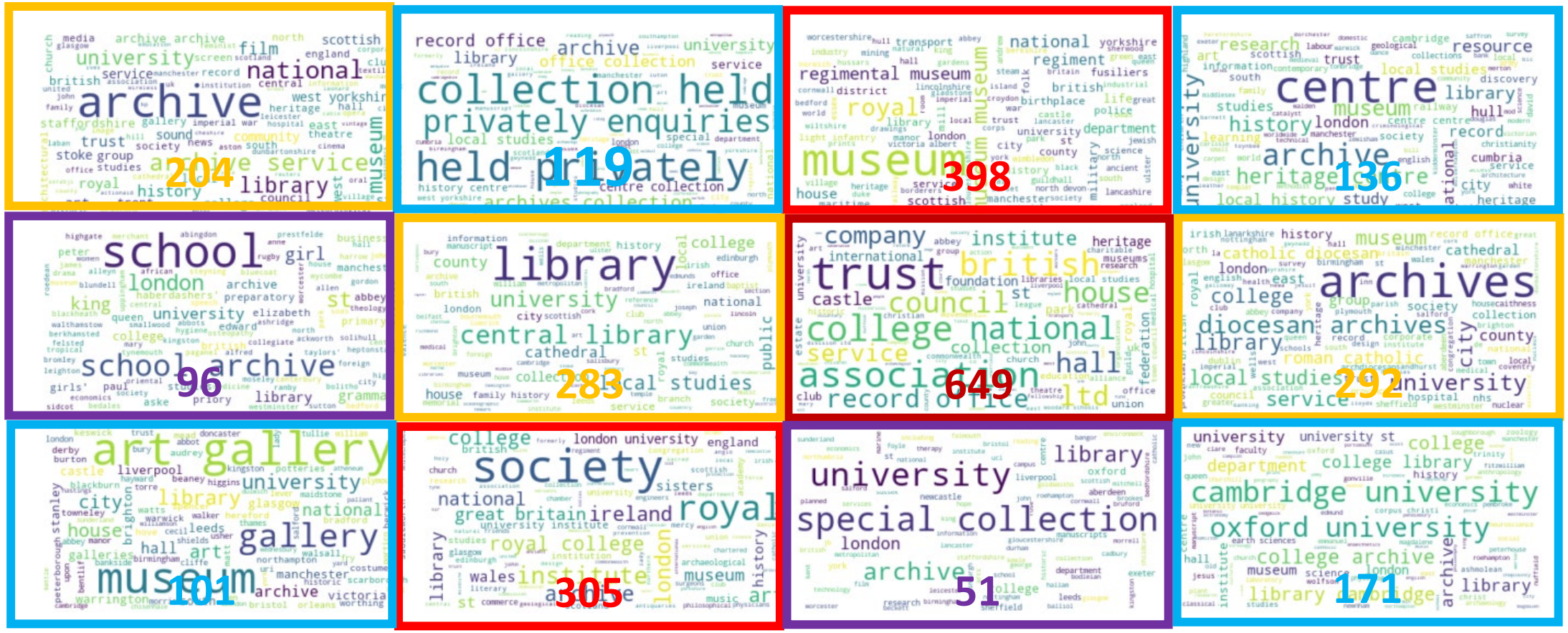


Image source: https://socraticlibrarian.files.wordpress.com/2011/02/library-cartoon.gif?w=557&h=386

Archives in the UK & RI Clustered

Over 3500 listed at discovery.nationalarchives.gov.uk – England 2309, Scotland 287, Wales 92, Republic of Ireland 88, Northern Ireland 29.



Data Challenge

- Data held in the archives are growing both in volume and complexity. Archive data suffer from data challenges (e.g. silo, lack of data, quality).
- Most solutions for privacy from the AI point of view tends to assume data as a something like structured data. Much of archive data sets are not statistical databases but things like letters, emails, images, sound, video.
- In many cases, the public and researchers request to see the raw data. In the arts and humanities many researchers combine close and distant reading for analyses – discombobulating approaches like differential privacy.
- Even where the sensitive nature of documents are understood, still onsite access adopted as a solution

Our Case Study: Email Archives

A gap in the debate about access and privacy

The Case Study

Arose due to considerations surround access to a born digital archive

- Born digital archive of an *avant garde* filmmaker
- The data: forensic disk images, emulations, extracted data
- Mixed personal and professional content
- Issue of access for different stakeholders – archivists, researchers, the public
- Difficulties arise from:
  - Volume
  - Variety
  - Noise
  - Privacy
- Email data encapsulated the issue

Stakeholder needs for emails

Archival

- Myriad of sensitive and personal data
- Identify, review large quantities of data effectively and in a timely way
- Manage access to a level reflective of risk
- Ethical and legal dimensions – tendency to err on the side of caution

Humanities researchers

- Wanting direct access for close examination
- Details of data are key for research
- Support to access unusual format and large scale

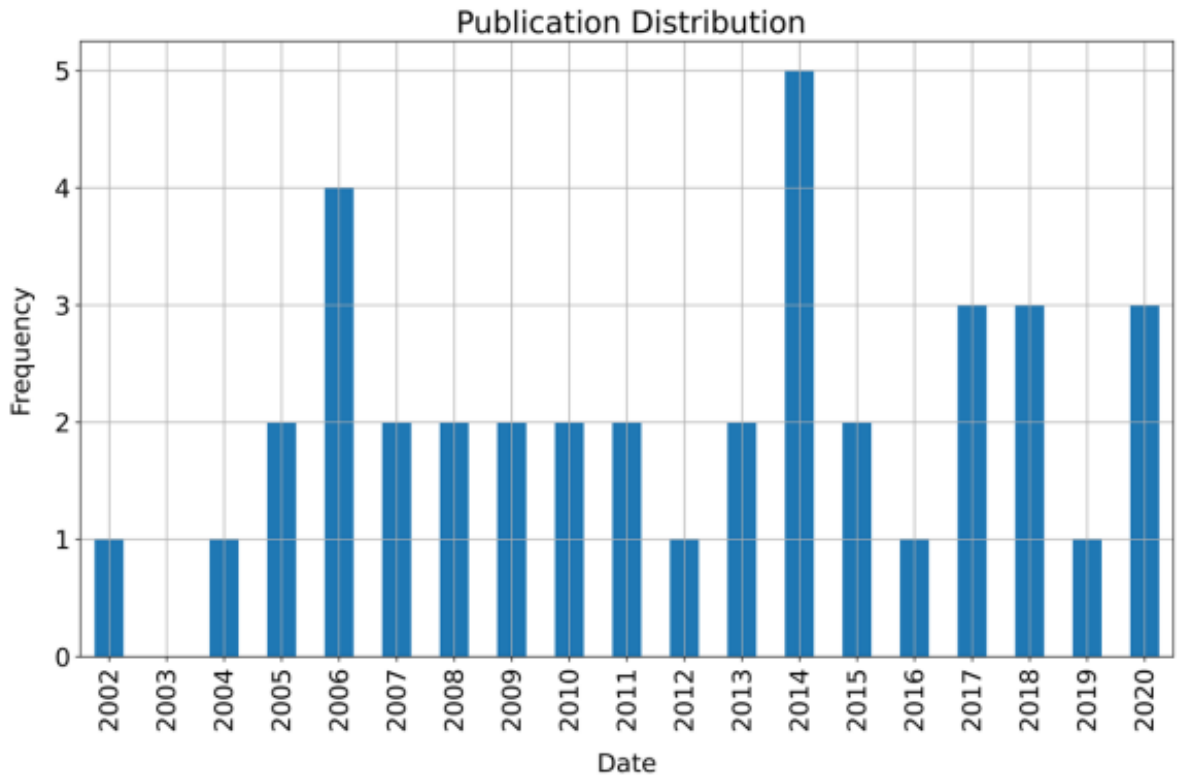
A visual solution?

- Visualisations have long been used to communicate complex data
- Visualisations have also been used to filter data, support anonymisation and decontextualisation as privacy management strategies (e.g. Chou et al. 2019)
- To this end, we set out to explore:
  - The research context and objectives of visualisation-based approaches to email research.
  - The levels of protection that can be offered by privacy aware strategies.
  - The potential impact of privacy management on the usefulness of the collection for humanities research.

First a review…

Landscape of the situation (Bartliff et al. 2022)

- “email~analysis” or “email~visualis(z)ation” or “email collection” AND archives, digital archives, or humanities AND privacy preserving, privacy management, privacy protection or scales of privacy.
- 39 papers surveyed
- 69 email visualisations



The number of articles included in the study (y-axis) by year of publication (x-axis)

Research priorities in email analysis

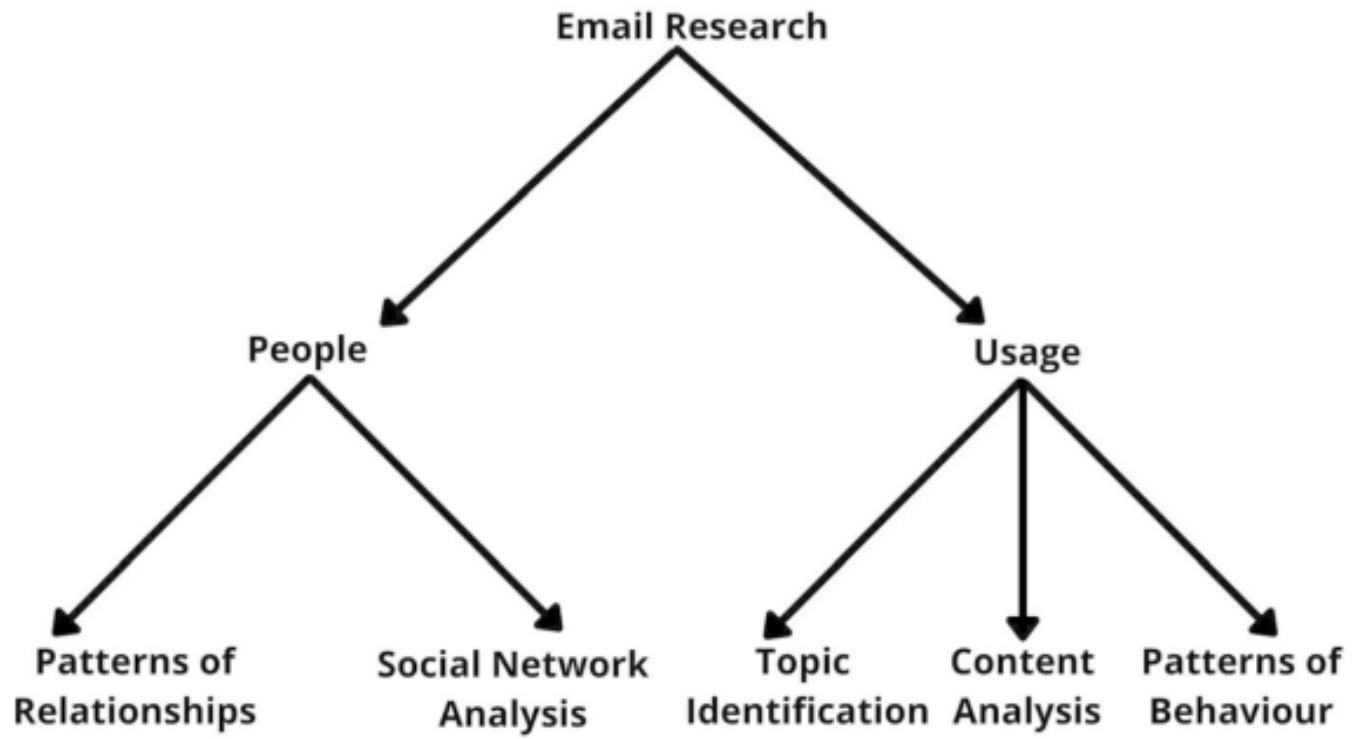


Fig. 2 A visual representation of the dyadic categorisation of email analysis and the more nuanced categories that sit within this

Distribution of visualisations across research priority

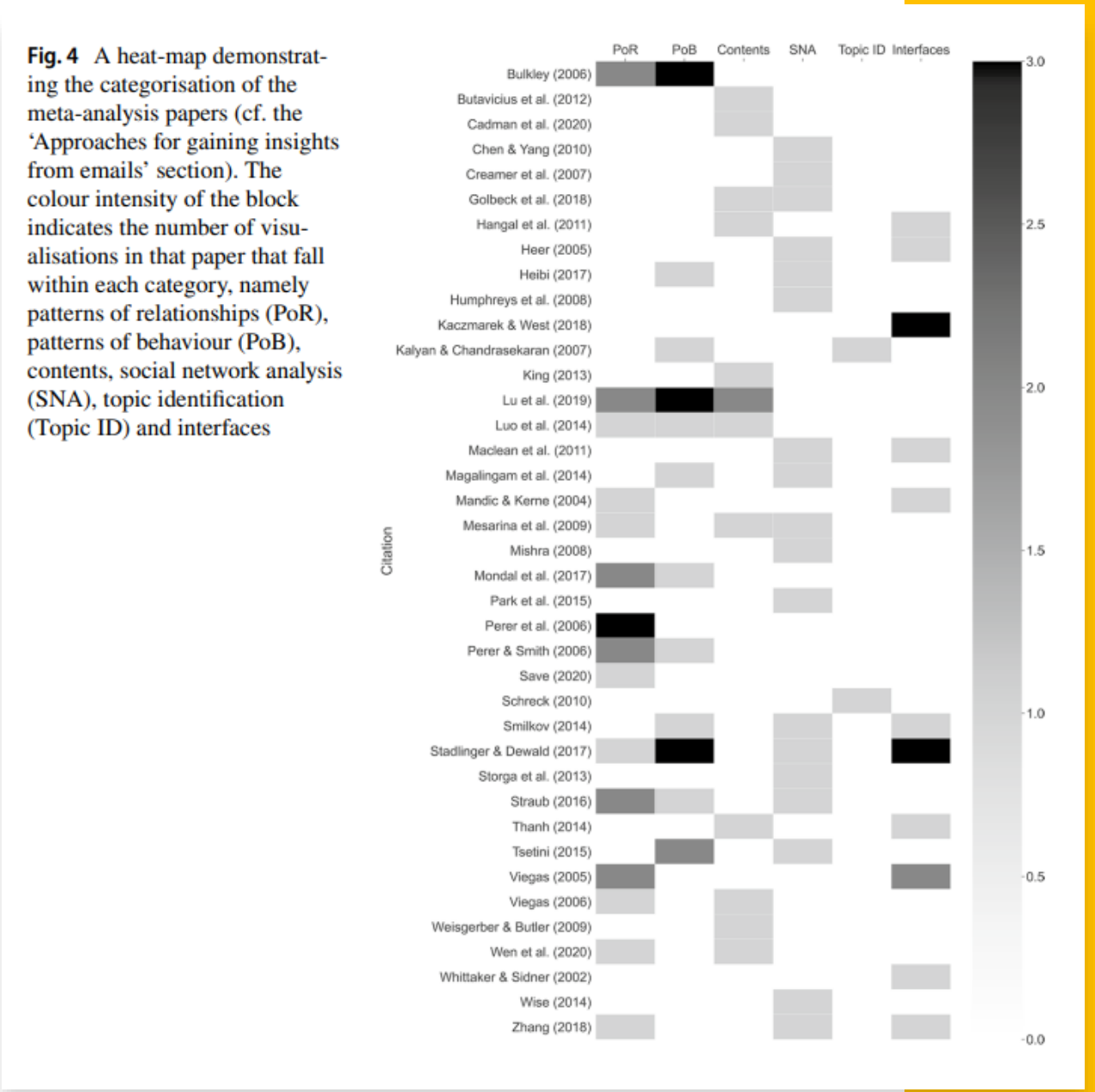


Fig. 4 A heat-map demonstrating the categorisation of the meta-analysis papers (cf. the 'Approaches for gaining insights from emails' section). The colour intensity of the block indicates the number of visualisations in that paper that fall within each category, namely patterns of relationships (PoR), patterns of behaviour (PoB), contents, social network analysis (SNA), topic identification (Topic ID) and interfaces

A scale of privacy management strategies

PrivCon0 - is reflective of data that is either accessed in its native environment or has been supplied to researchers in a state that mirrors this, with minimal intervention.

PrivCon1 - includes situations whereby the data have been altered or removed in order to obscure the identity of individuals contained within.

PrivCon2 - involves the grouping or amalgamation of data to the point that individuals become' lost in the crowd', minimising the risk that details might be identified.

PrivCon3 – a form of privacy management involves shifting the data through the use of an algorithm, statistical model or encryption, in a way that maintains the statistical characteristics of the data-set but the detail does not consistently refect the original.

PrivCon4 - the antithesis of PrivCon0, refers to the practice of keeping an archive 'dark', inaccessible to researchers without special, often on site, permissions.

Distribution of privacy management strategies

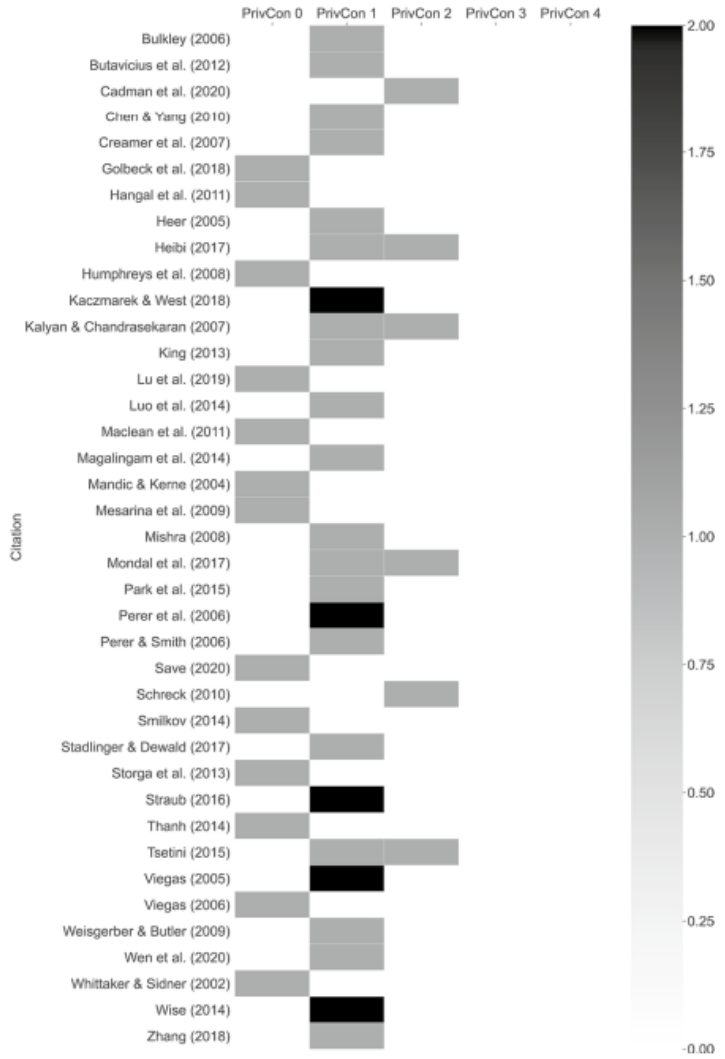


Fig. 5 A heat-map demonstrating the distribution of the reviewed papers across the PrivCon scale. The colour intensity of the block indicates the number of papers that fall within each category, namely patterns of relationships (PoR), patterns of behaviour (PoB), contents, social network analysis (SNA), topic identification (Topic ID) and interfaces

Fig. 6 A bar chart showing the application of different privacy aware strategies (y-axis) and how these are distributed across the categorisations of research interests (x-axis)

Privacy management by research priority

Summary of PrivCon levels

Level	Papers	Example Data	Pros	Con
PrivCon 0	13	forensic evidence one's own email data	low labour full access by user	high risk
PrivCon 1	30	Enron email datasets for ML	some protection automated tools exist details accessible	identity could be inferred reliability can vary
PrivCon 2	6	Enron curated commercial datasets	can track flow, activity trends accessible	aggregated stats only multiple searches can reveal
PrivCon 3	0	-	secure in theory	difficult to implement

Key discussion points

- In A&H, researchers often utilise close or manual methods.
- Increasing the stringency of privacy management decreases useful access.
- This conflicts with Archival priorities.
- Only half of the papers (19/39) make explicit mention of privacy – an inconvenience?
- 9/39 focus on tester's own data. 8/39 on public data (e.g. enron)
- Only 2 engage with personal digital archives – for one, a co-author is the owner of the archive. 'not a luxury we expect most historians and social scientists to have' (Perer 2006)
- Visualisation might leverage the flexibility of the digital format, promote exploratory search behaviour for both holistic and specific observations BUT
- Visualisation is also known to reveal previously unseen patterns that might jeopardise privacy.
- **The data is not yet there to determine efficacy of visual methods for privacy management OR researcher access.**

Towards privacy-aware exploration of archived personal emails

Progressing the debate about access and privacy

A rare opportunity

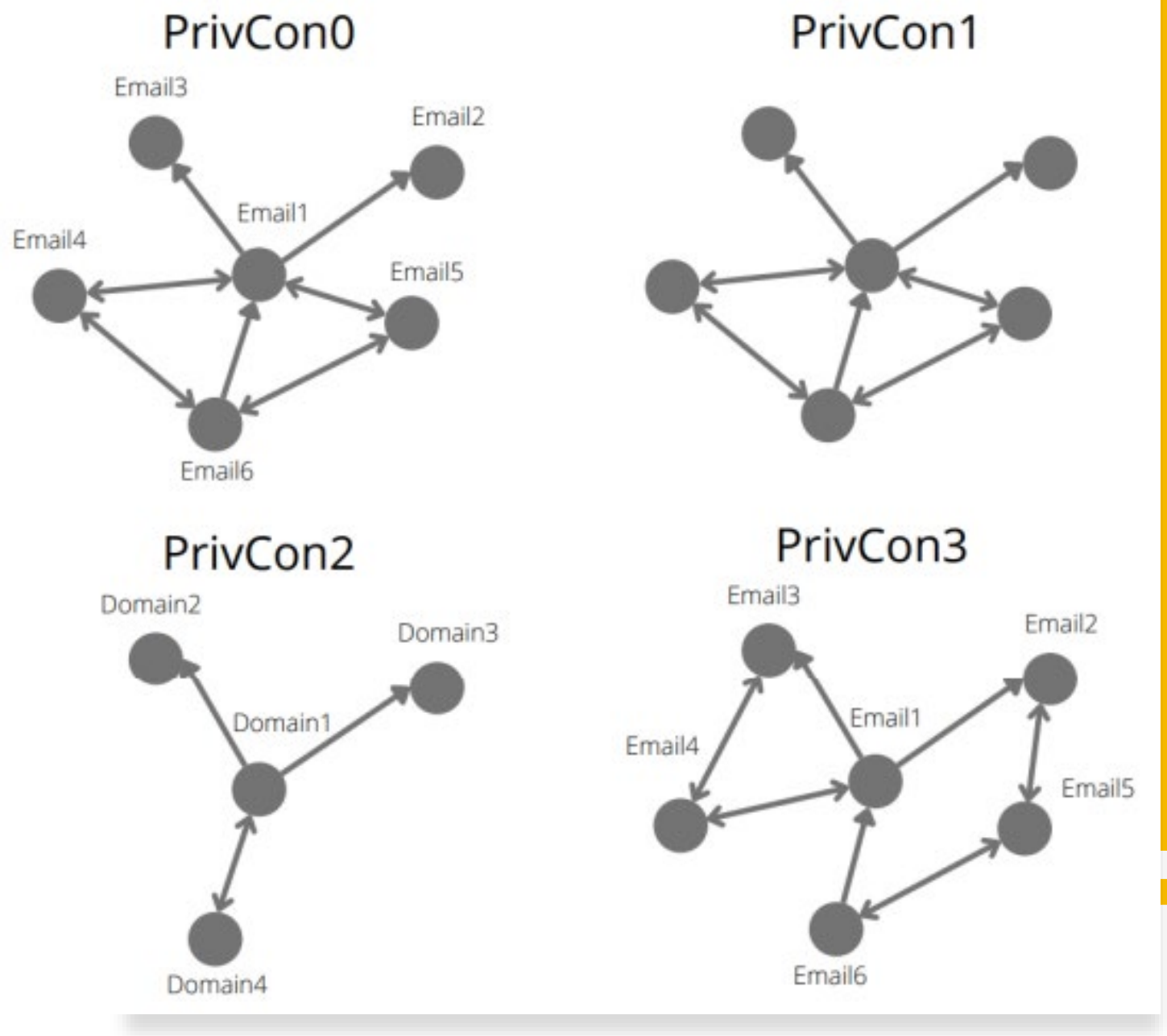
- A project focused on exploring a personal hybrid archive
- Privileged access to archival content, and a project team of data scientists, experienced archivists, practicing artists, and arts & humanities researchers.
- A ready-made user test group to explore the impact and efficacy of privacy conscious visualisation strategies.
- **Paper to be published in coming months in IJDL**

The experiment design

- Delphi method* - experts answer questionnaires in two or more rounds with access to summary of the previous round.
- RQ1: What is the relationship between the extent of privacy-awareness applied to visualisations of email collections and the usefulness of these visualisations to researchers/practitioners?
- RQ2: What design features of the privacy aware visualisations are the most/least useful for researchers and practitioners as an interface for the email collection?
- Testing five types of visualisation at a range of PrivCon levels
- User group is composed of two archivists and two A&H researchers each with different priorities in their work

*https://en.wikipedia.org/wiki/Delphi_method

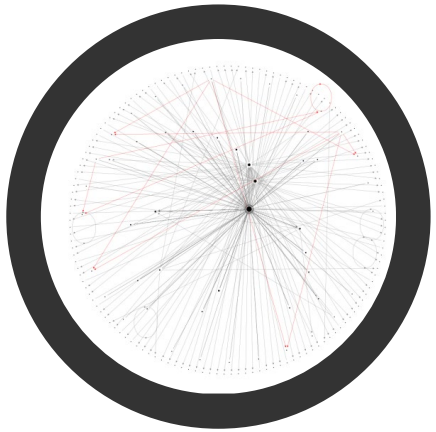
Visual representation of the PrivCon levels*

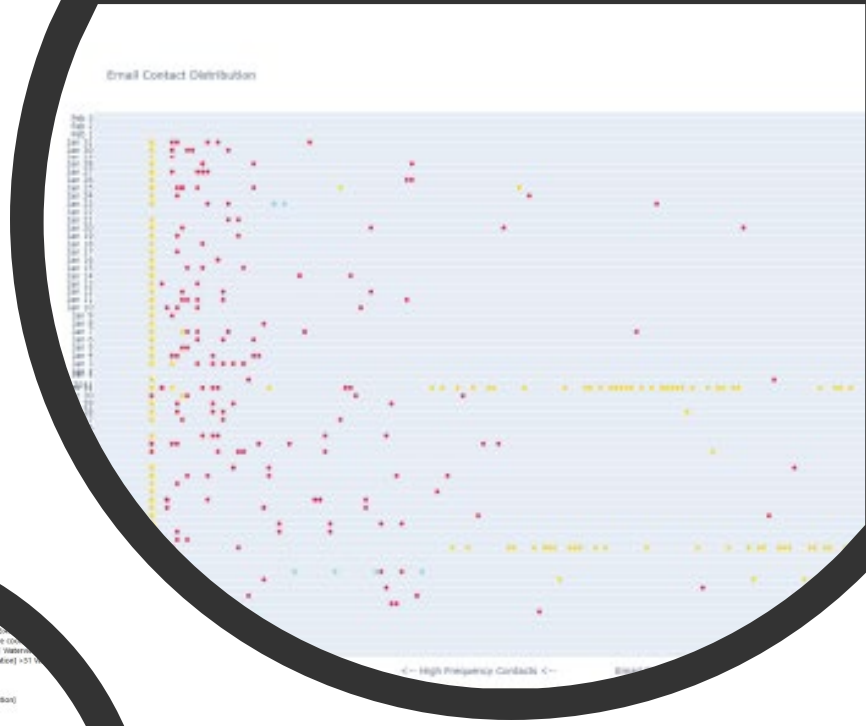


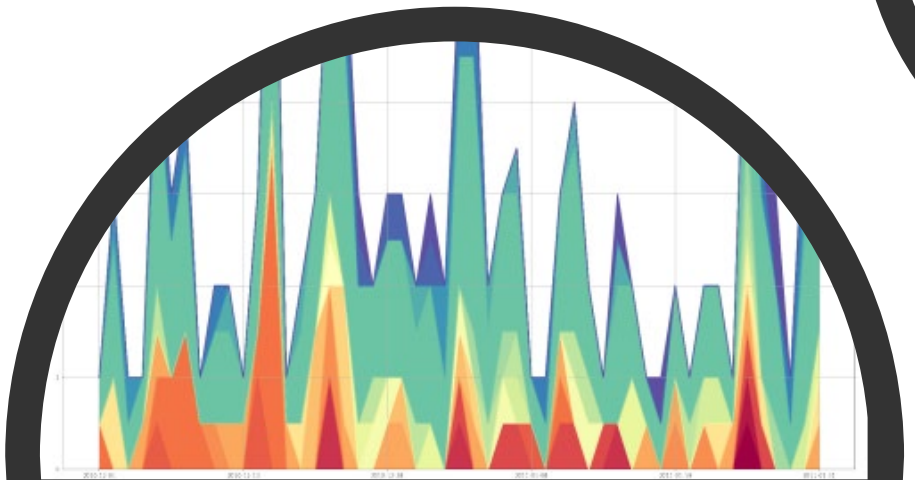
*Introduced in slide 21

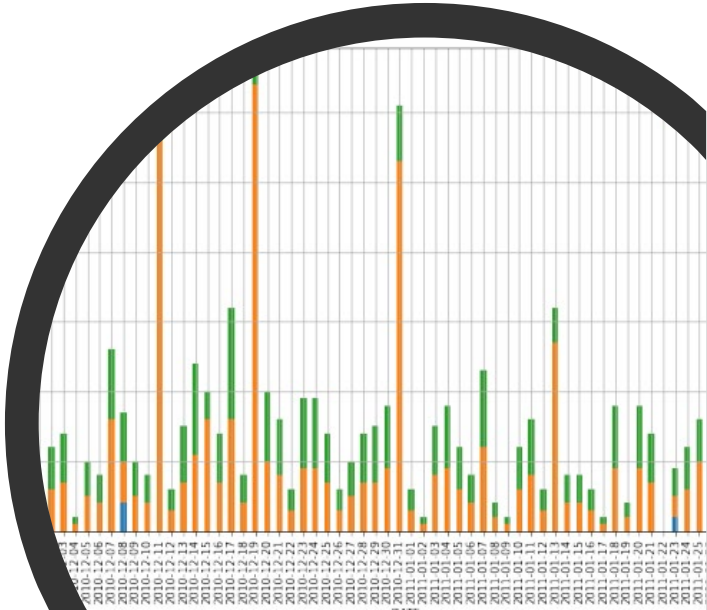
Chosen visualisation types

- Network graphs – typical of Social Network Analysis focused research (created at PrivCon 0-3)
- Mountain graphs (stacked line graphs) – ebb and flow of interpersonal relationships over time (created at PrivCon 0-3)
- Scatter plots – points of contact for relationships (created at PrivCon 0-1)
- Bar Graphs – demonstrating patterns of behaviour, in this case use of To, From, CC and BCC (created at PrivCon 0-3)
- Word Trees – email content. Usually dynamic, but static for the study (created at PrivCon 0-2)









Stages of the Study (a)

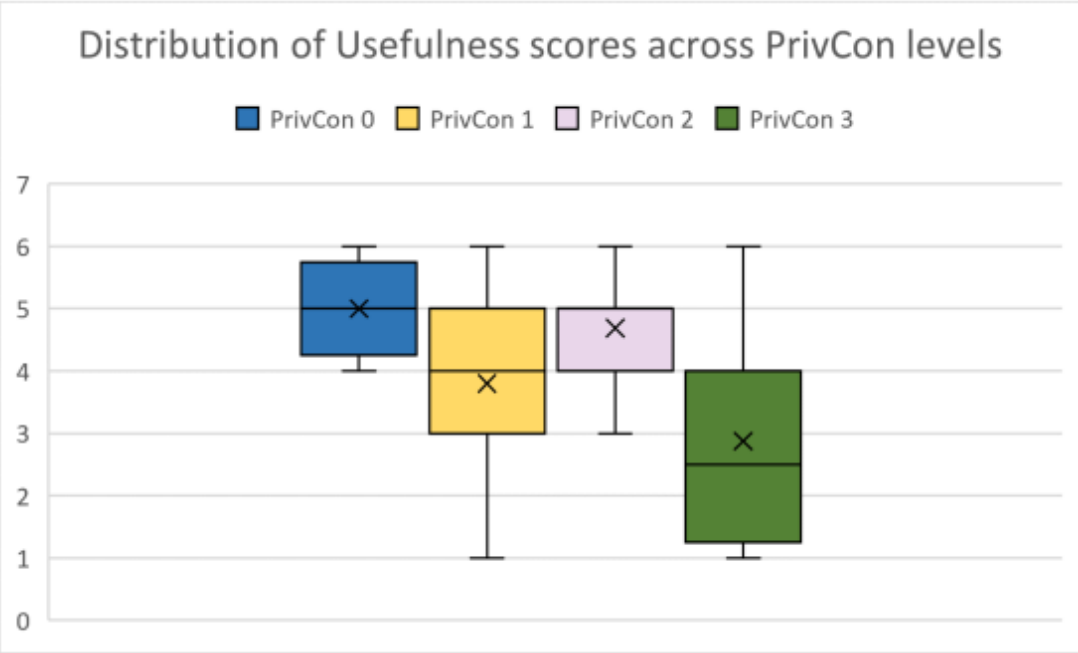
- Stage one – Contextualising the participants' background, research, and/or practice.
- Stage two –
  - Summary of stage 1 responses, opportunity to review own response
  - Presented with the visualisations (grouped by type but with visualisation type and level of privacy ordered randomly)
  - For each visualisation, participants were asked:
    1. What kinds of information can you gather from this visualisation?
    2. Does this type of visualisation support your approaches to research?
    3. In what ways might visualisations like this help you to address your key questions/themes and/or envisioned outcomes?
    4. In what ways could the visualisation be lacking in helping you address your key questions/themes and/or envisioned outcomes?

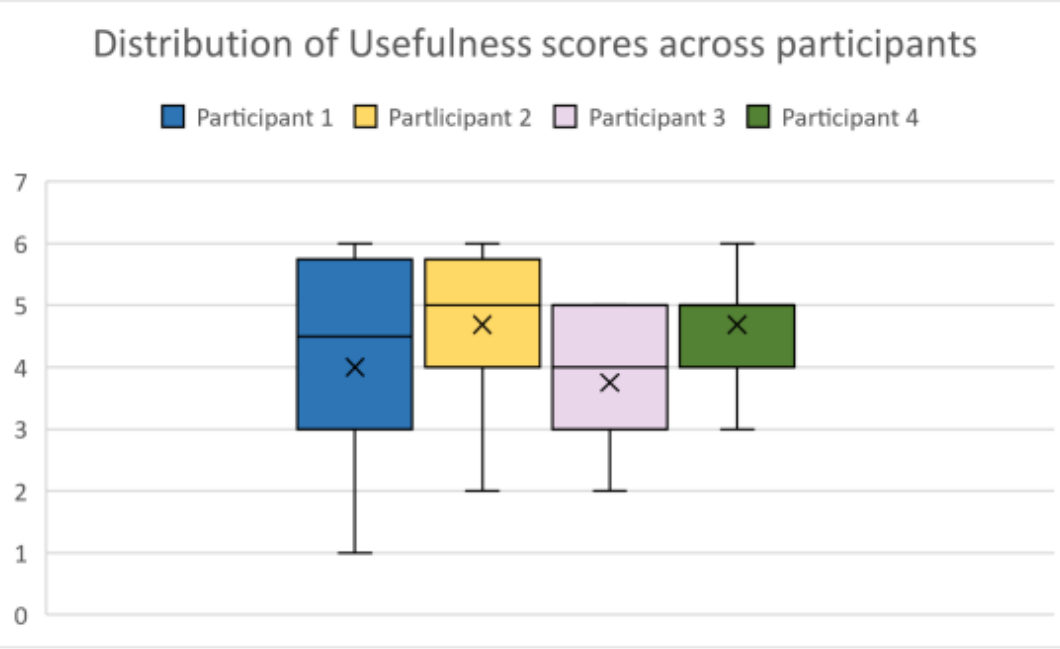
Stages of the Study (b)

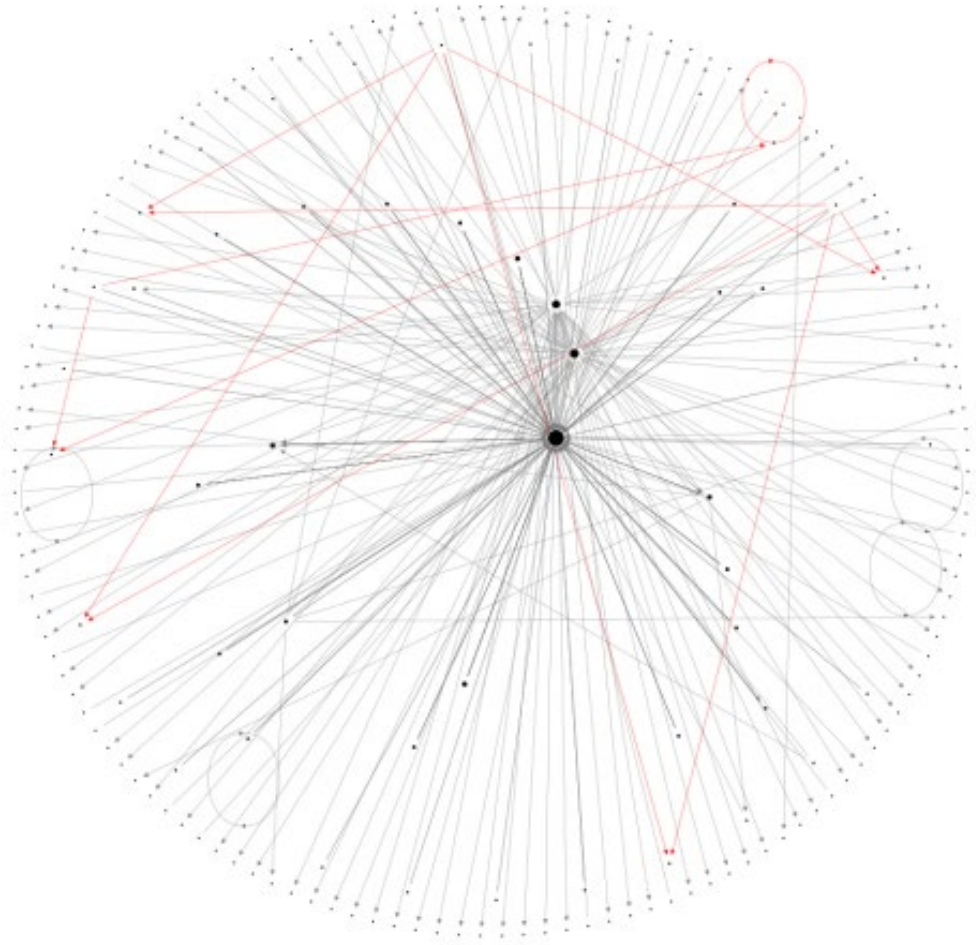
- Stage three –
  - Summary of stage 2 responses, opportunity to comment on own response
  - Ranking of visualisation based on usefulness to their work
  - Description of reasoning for the rank
  - Specifically
    1. Is there anything you would like to add or change in relation to your initial assessment of this visualisation?
    2. How useful is this visualisation for your research or practice? 1 (not useful) - 7 (very useful)
    3. Why have you given this rating?

Findings regarding usefulness

	No. of Responses	Average score
Word Tree	12	4.583
Directed Network Graph	16	3.688
Mountain Graphs	12	5.000
Scatter Graphs	8	4.125
Bar Charts	16	4.188

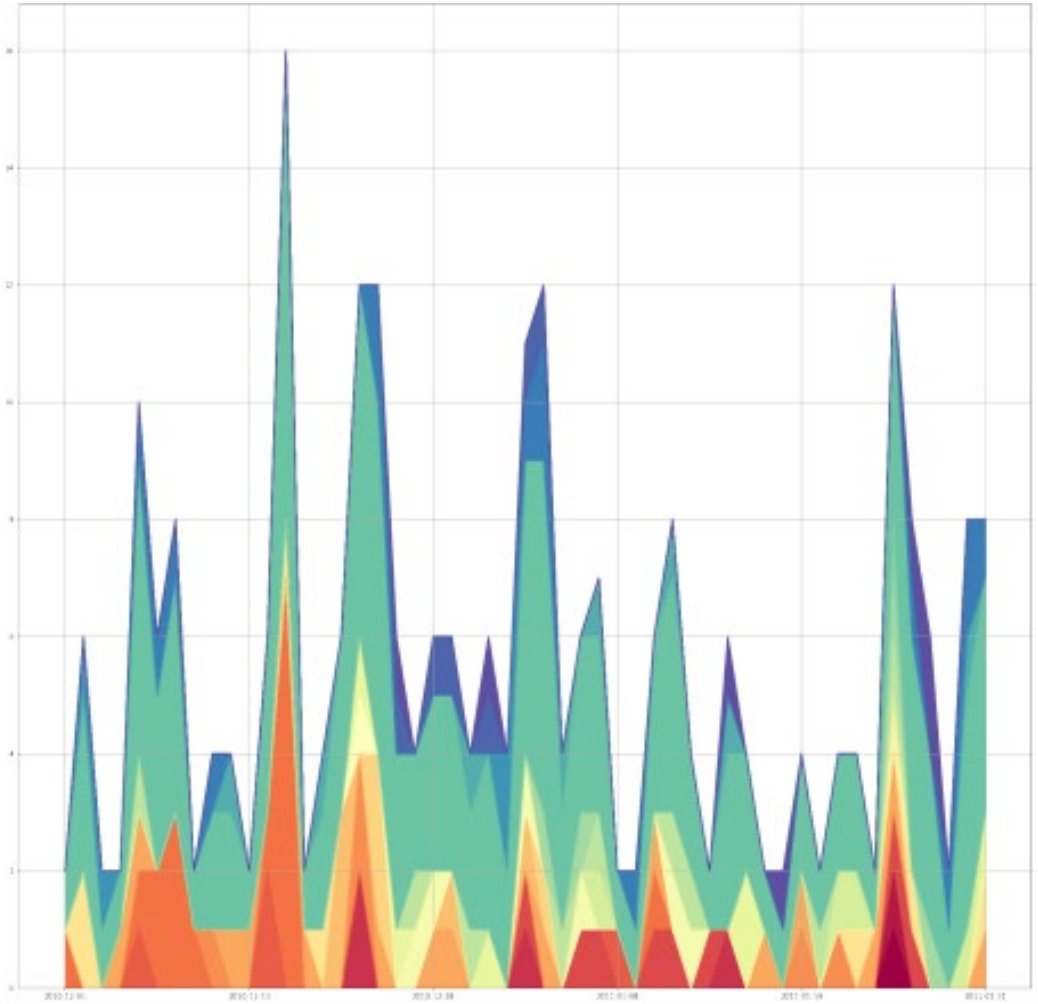






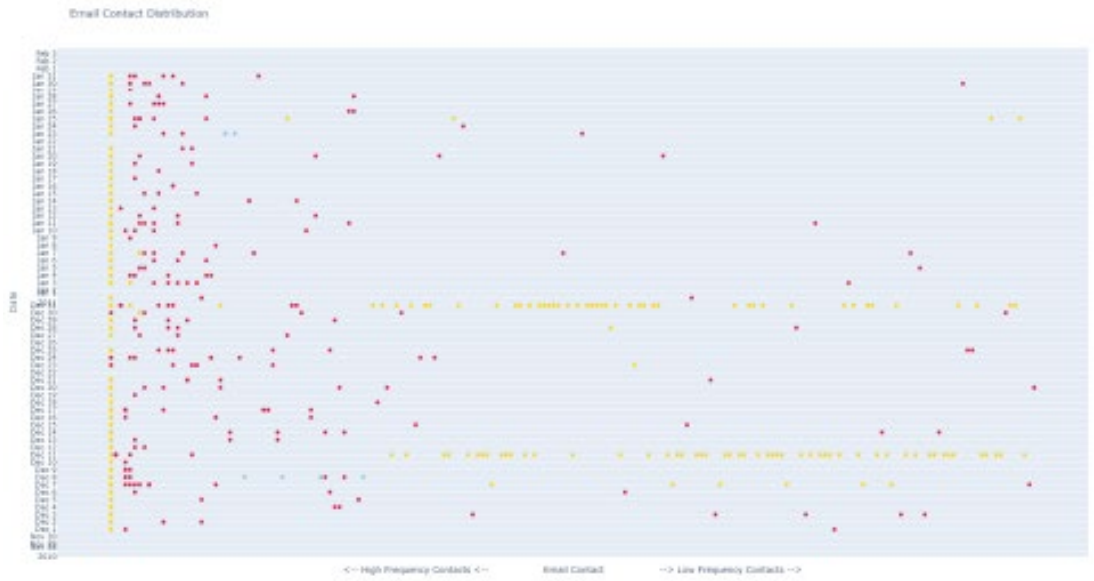
PrivCon Scale	P1	P2	P3	P4	Average	Standard Deviation
0	5	5	4	5	4.75	0.433
1	3	2	2	4	2.75	0.829
2	5	5	4	6	5	0.707
3	1	2	2	4	2.25	1.090

Details for network graphs



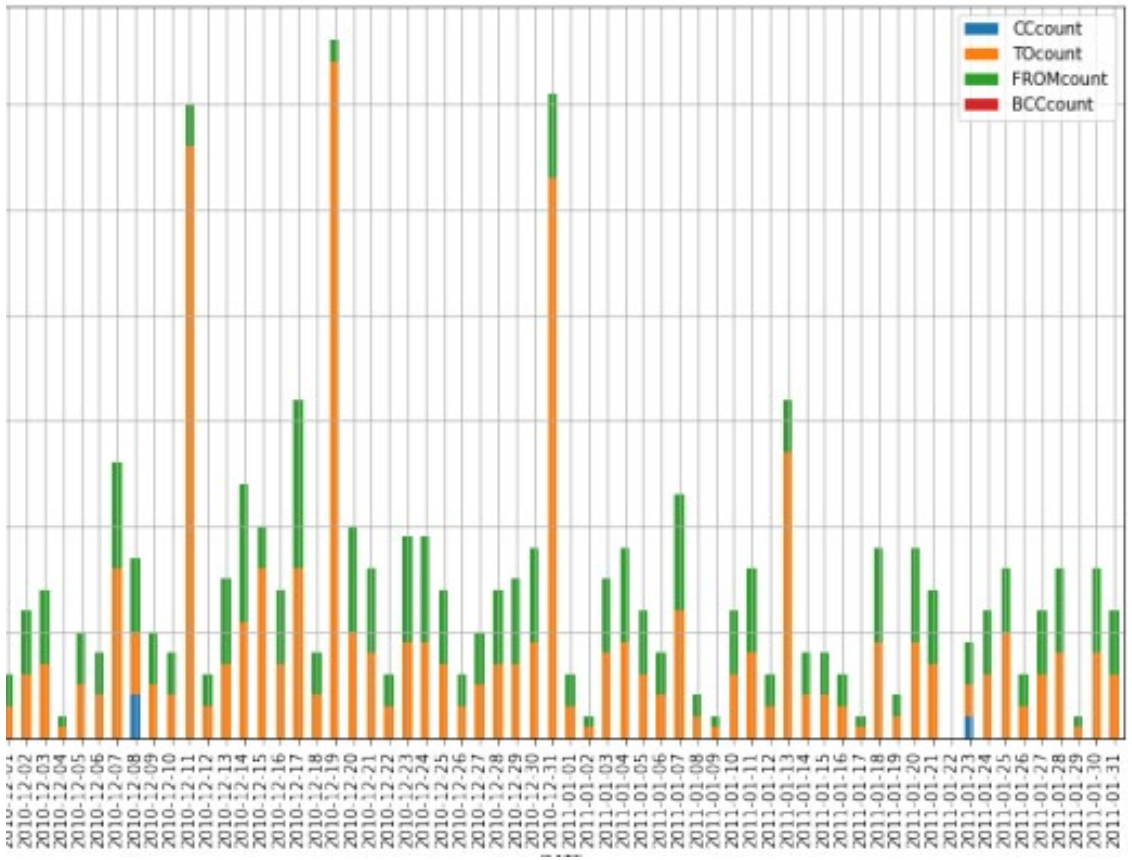
PrivCon Level	P1	P2	P3	P4	Average	Standard Deviation
0	4	5	5	5	4.75	0.433
1	6	5	4	5	5	0.707
2	6	5	5	5	5.25	0.433

Details for mountain graphs



PrivCon Level	P1	P2	P3	P4	Average	Standard Deviation
0	6	4	4	6	5	1.000
1	3	4	3	3	3.25	0.433

Details for scatter plots



PrivCon Level	P1	P2	P3	P4	Average	Standard Deviation
0	6	6	5	4	5.25	0.829
1	1	6	3	4	3.5	1.803
2	4	6	3	5	4.5	1.118
3	1	6	3	4	3.5	1.803

Details for bar charts

PrivCon Level	P1	P2	P3	P4	Average	Standard Deviation
0	5	5	5	6	5.25	0.433
1	3	5	5	5	4.5	0.866
2	5	4	3	4	4	0.707

Details for word trees

Discussion

- Participants were able to engage creatively and productively with the majority of the visualisations. Sometimes concomitant with existing research/practice methods, other times new perspectives.
- Contradicts theorised inverse relationship between privacy level and usefulness
- Rather usefulness is dependent on the underlying focus of the data and associated analysis
- PrivCon1 viewed as removing essential information – overarching pattern insufficient for their work. BUT safer to release archival data.
- PrivCon2 striking results – slightly less useful that PrivCon0 BUT greater range of opportunities to engage with data – new patterns to provoke thinking and understanding of data
- PrivCon3 least useful – potential knowledge gap in understanding, resulting in anxiety from participants about their understanding of the visualisation

Conclusions

Filling the gap in the debate between AI, access and privacy

Key points

- Privacy goes beyond legal and ethical codes and can be a highly personal concept
- The exponential growth of digital content and an increasingly networked world challenge open access due to privacy.
- AI and machine learning use our data – privacy as an increasing concern. Not all solutions for AI help situation in archives.
- Visualisations, and the techniques that underly their creation, offer a potential gradated approach to privacy management reflective of stakeholder needs.
- Study has direct implications for archives and their strategies for managing digital content to allow timely access.
- Also, broader implications for modern data explorers.
- Only the beginning for addressing the debate.

The Future: demand for a privacy engineer

- Specialisation in data privacy will be across industry, academia, public sectors as a key requirement for ethics boards for business, research and governance.
- Critical in finance or healthcare tightly governed by a number of regulatory frameworks about data protection, sensitive information and privacy.
- Customer engagement and expansion by demonstrating you know how to manage sensitive data.
- Core element in product design and solutions that incorporate attention to data privacy.
- Developing responsible AI reflecting privacy and human rights and associated risks – how would you implement it, evaluate it, measure it, communicate it?
- See key themes of shortlisted RAI UK Keystone projects and the BRAID UK blog on AI safety.

References

- Bartliff, Z., Kim, Y. & Hopfgartner, F. A survey on email visualisation research to address the conflict between privacy and access. Arch Sci 22, 345–366 (2022). https://doi.org/10.1007/s10502-022-09387-2
- Chou J-K, Wang Y, Ma K-L (2019) Privacy preserving visualization: a study on event sequence data. in 'Comput Gr Forum', Vol 38, Wiley Online Library, pp 340–355
- Perer A, Shneiderman B, Oard DW (2006) Using rhythms of relationships to understand e-mail archives. J Am Soc for Inf Sci Tech 57(14):1936–1948