

# Leveraging mailing list archives for digital history research: A case study on mining the Humanist discussion group

Marco Humbel, Jiajie Zhang, Valentina Prishchepova, Aleksey  
Varfolomeyev, Julianne Nyhan and Andreas Vlachidis

Mixed Methods Digital Oral History  
2025



# Scope

- Mailing list archives as primary source material
- What are the potentials of mailing list archives for exploring the formation of academic disciplines?
- Case study: Humanist Discussion Group (est. 1987)

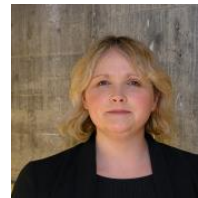
# Project: Mixed-methods Digital Oral History

Enfolding semantic web technologies and historical-interpretative analysis

Andrew Flinn  
Associate Professor  
UCL Information Studies



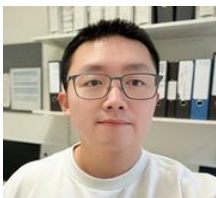
Julianne Nyhan  
Professor  
TU Darmstadt Institute of History



Andreas Vlachidis  
Associate Professor  
UCL Information Studies



Jiajie Zhang  
Research Associate  
UCL Information Studies



Frieda Schmidt  
Research Associate  
TU Darmstadt Institute of History



Valentina Prishchepova  
Student Assistant  
TU Darmstadt Institute of History



Marco Humbel  
Research Associate  
UCL Information Studies  
TU Darmstadt Institute of History



Florian Elleringmann  
Student Assistant  
TU Darmstadt Institute of History



Aleksey Varfolomeyev  
Technical Assistant  
(Multimodal Digital Oral History Project)  
TU Darmstadt Institute of History



# The Humanist: Access and Data preparation

- Goal:
  - Identify influential people, conferences and institutions
  - Detect thematic and disciplinary shifts over time
- 3 archives:
  - 1987-2008
  - 2008-2018
  - 2018-present
- UCL Ethics approval (AH/2025/23)



## Index of /archives/Converted\_Text/

../		
<a href="#">humanist.1987-1988.txt</a>	26-Oct-2018 11:09	2354980
<a href="#">humanist.1988-1989.txt</a>	26-Oct-2018 11:09	1535924
<a href="#">humanist.1989-1990.txt</a>	26-Oct-2018 11:09	4663437
<a href="#">humanist.1990-1991.txt</a>	26-Oct-2018 11:09	4973851
<a href="#">humanist.1991-1992.txt</a>	26-Oct-2018 11:09	3243062
<a href="#">humanist.1992-1993.txt</a>	26-Oct-2018 11:09	2934777
<a href="#">humanist.1993-1994.txt</a>	26-Oct-2018 11:09	2776195
<a href="#">humanist.1994-1995.txt</a>	26-Oct-2018 11:09	1739016
<a href="#">humanist.1995-1996.txt</a>	26-Oct-2018 11:09	2494851
<a href="#">humanist.1996-1997.txt</a>	26-Oct-2018 11:09	2109480
<a href="#">humanist.1997-1998.txt</a>	26-Oct-2018 11:09	3678774
<a href="#">humanist.1998-1999.txt</a>	26-Oct-2018 11:09	2907550
<a href="#">humanist.1999-2000.txt</a>	26-Oct-2018 11:09	2652849
<a href="#">humanist.2000-2001.txt</a>	26-Oct-2018 11:09	3232307
<a href="#">humanist.2001-2002.txt</a>	26-Oct-2018 11:09	2634930
<a href="#">humanist.2002-2003.txt</a>	26-Oct-2018 11:09	2419876
<a href="#">humanist.2003-2004.txt</a>	26-Oct-2018 11:09	2834023
<a href="#">humanist.2004-2005.txt</a>	26-Oct-2018 11:09	2718227
<a href="#">humanist.2005-2006.txt</a>	26-Oct-2018 11:09	2665694
<a href="#">humanist.2006-2007.txt</a>	26-Oct-2018 11:09	2238595
<a href="#">humanist.2007-2008.txt</a>	26-Oct-2018 11:09	1868483

## Images:

Humanist logo: <https://dhhumanist.org/>

Humanist archive: [https://dhhumanist.org/archives/Converted\\_Text/](https://dhhumanist.org/archives/Converted_Text/)

# The Humanist: Method

- Data acquisition
  - File download
  - Webscraping with Wget
- Data cleaning
  - Regular Expressions
  - OpenRefine
- Data processing and analysis
  - spaCy
  - Matplot

```
From humanist-bounces@lists.digitalhumanities.org Thu May 7 08:40:39 2009
Return-Path: <humanist-bounces@lists.digitalhumanities.org>
X-Original-To: humanist-archiver@digitalhumanities.org
Delivered-To: humanist-archiver-digitalhumanities@woodward.joyent.us
Received: from woodward.joyent.us (localhost [127.0.0.1])
    by woodward.joyent.us (Postfix) with ESMTP id D5A136821;
    Thu, 7 May 2009 08:40:39 +0000 (GMT)
Received: by woodward.joyent.us (Postfix, from userid 1006)
    id B19E1680B; Thu, 7 May 2009 08:40:37 +0000 (GMT)
From: Humanist Discussion Group <willard.mccarty@mccarty.org.uk>
To: humanist@lists.digitalhumanities.org
MIME-Version: 1.0
Message-Id: <20090507084037.B19E1680B@woodward.joyent.us>
Date: Thu, 7 May 2009 08:40:37 +0000 (GMT)
Subject: [Humanist] 23.2 invitation to join the iSchools
X-BeenThere: humanist@lists.digitalhumanities.org
X-Mailman-Version: 2.1.10
Precedence: list
Reply-To: Online seminar for digital humanities
    <humanist@lists.digitalhumanities.org>
Content-Type: text/plain; charset="us-ascii"
Content-Transfer-Encoding: 7bit
Sender: humanist-bounces@lists.digitalhumanities.org
Errors-To: humanist-bounces@lists.digitalhumanities.org
```

Custom text transform on column Date

Expression Language General Refine Expression Language (GREL) ▼

`value.toDate("EEE, d MMM yyyy HH:mm:ss Z").toString("yyyy-MM-dd")` No syntax error.

Preview History Starred Help

row	value	value.toDate("EEE, d MMM yyyy ...
1.	Wed, 13 Feb 2008 14:18:59 +0000	2008-02-13
2.	Wed, 13 Feb 2008 14:19:44 +0000	2008-02-13
3.	Thu, 14 Feb 2008 13:39:59 +0000	2008-02-14
4.	Thu, 14 Feb 2008 13:40:58 +0000	2008-02-14
5.	Thu, 14 Feb 2008 13:44:36 +0000	2008-02-14
6.	Thu, 14 Feb 2008 13:43:51 +0000	2008-02-14

On error ☒ keep original ☐ set to blank ☐ store error ☐ Re-transform up to  times until no change

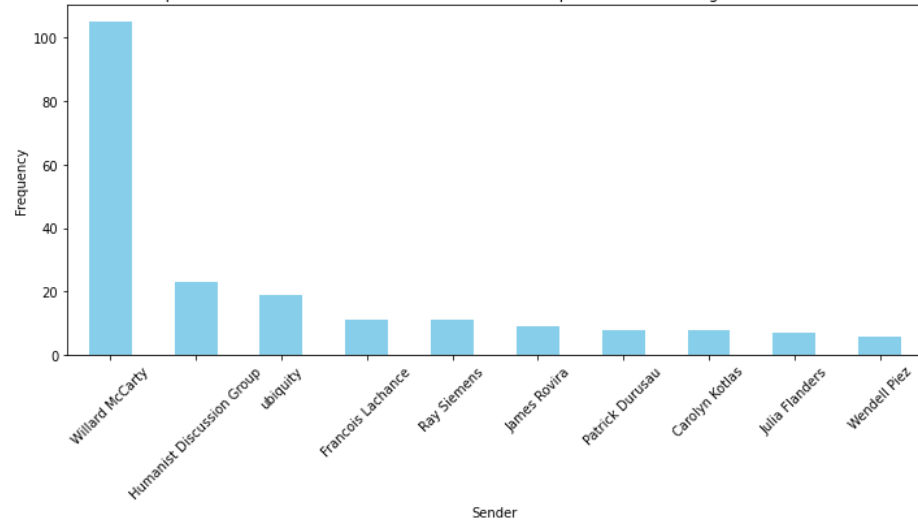
OK Cancel

## Images:

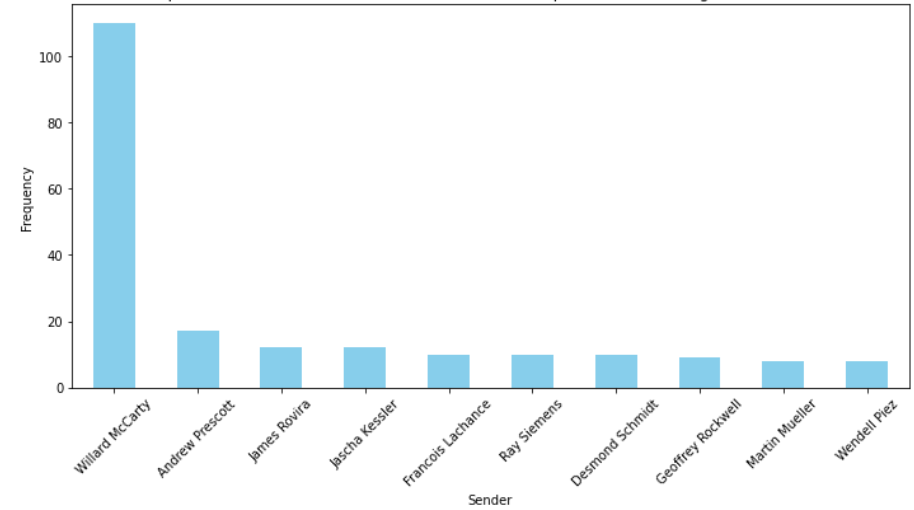
Metadata screenshot: <https://dhhumanist.org/archives/Current/Humanist.vol23.txt>

# The Humanist: Initial results

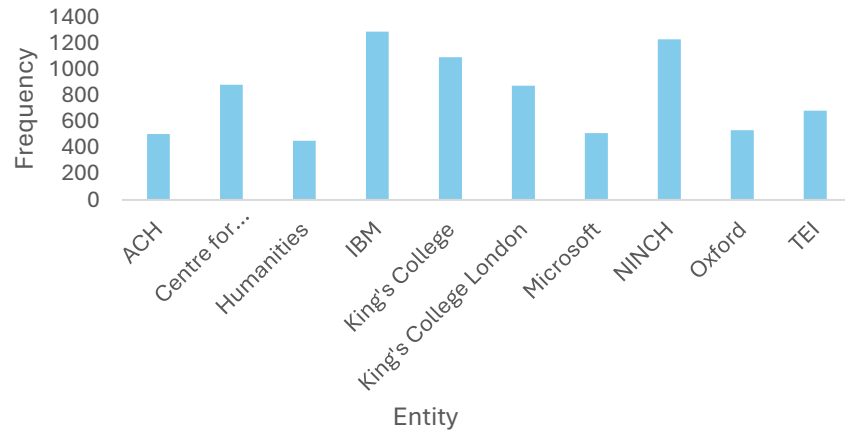
Top 10 senders to the Humanist from a random sample of 4,511 messages (2005-2010)



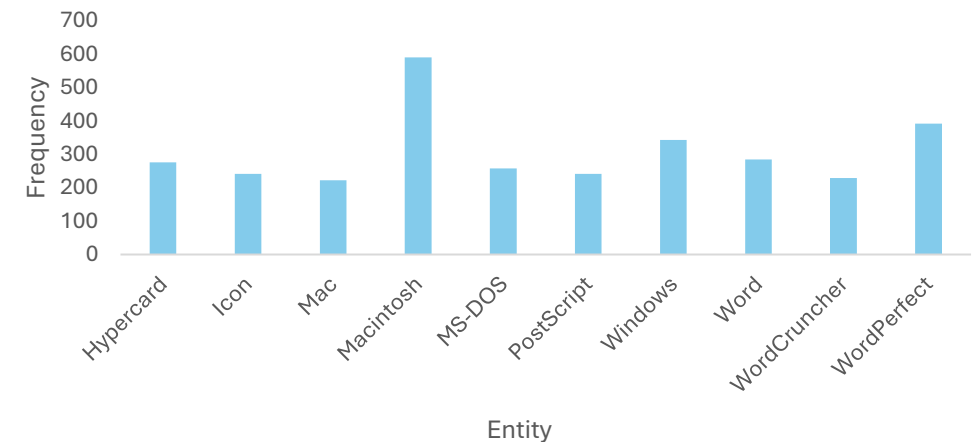
Top 10 senders to the Humanist from a random sample of 4,511 messages (2010-2015)



Top 10 most frequent organisations mentioned (whole corpus)



Top 10 most frequent products mentioned (whole corpus)



# Project: Next steps



Moderate expectations, tolerable disappointments



It's probably the only modestly widely used system with a command language in Latin



They took a chance: Susan Hockey and Julianne Nyhan



It's a little mind-! Agüera and Julia

## Interview

JN

Please reflect on your earliest memories of encountering a computer or computing technologies.

SH

Well, I'll tell you how I got started. I was an undergraduate in the late 1960s. I did Classics at Oxford and then did my final degree in Egyptian with Akkadian. I was always interested in language things, and I think it was in about 1967 that I remembered reading those articles in the *Observer* from Andrew Morton (see, for example, Morton 1963), who'd been doing this text analysis study of the Pauline Epistles with a computer (Morton 1965). It sounded really interesting and I thought I'd like to work in something like that. So, I checked how I could do this. I met one of the lecturers in Arabic, called Alan Jones, and found out that he was already doing some computing things (see Jones 1971). I think my tutor told me about him and I met him and found out that he was doing some text analysis work on the Koran. Because technology within the universities at that time was quite small, and very much focussed on Sciences, he was doing his work at an organisation called the Atlas Computer Laboratory,<sup>3</sup> which was funded by what was then called the Science Research Council to provide computing support for universities – the things that the universities couldn't have the technologies to do themselves.



## Images:

Screenshot archive: <https://hiddenhistories.net/s/medorah/page/OralHistory>

Interview: Nyhan, J., Flinn, A. (2016). They Took a Chance: Susan Hockey and Julianne Nyhan. In: Computation and the Humanities. Springer Series on Cultural Computing. Springer, Cham.

[https://doi.org/10.1007/978-3-319-20170-2\\_6](https://doi.org/10.1007/978-3-319-20170-2_6)

Preliminary network: Jiajie Zhang (2025)



Thank you!

marco.humbel.17@ucl.ac.uk