



DIGITAL HUMANITIES & RESEARCH SOFTWARE ENGINEERING SUMMER SCHOOL

EDINBURGH, JULY 02-05, 2024

IN COLLABORATION WITH:



The
Alan Turing
Institute



SPONSORED BY:



The
Alan Turing
Institute

TODAY'S SCHEDULE

The Alan Turing Institute

Welcome and Introduction (Federico Nanni)

Research Software Engineering in the Arts and Humanities: a community-driven approach (Stavros Angelis)

DH&RSE at Turing:

- Collaboration in Digital Humanities Research (Kaspar Beelen)
- NLP Research in LwM (Federico Nanni)
- DH & RSE practices: the Seshat Project (Ed Chalstrey)

Collaborative Coding activity

DH&RSE Career Talk (Kalle Westerling)

Housekeeping

- Toilets
- Food Consumption
- Water Fountains
- Fire Alarm
- Code of Conduct

Collaboration in Digital Humanities Research

Memoirs of a historian among the machines

DH & RSE Summer School

5th July, 2024

Kaspar Beelen

Technical Lead Digital Humanities, School of Advanced Study, University of London;
Senior Research Associate, The Alan Turing Institute

Overview of this talk

- Introduction: A Historian among the Machines
- Living with Machines: Collaboration at Scale
- Reflections: Pros and Cons of Collaboration



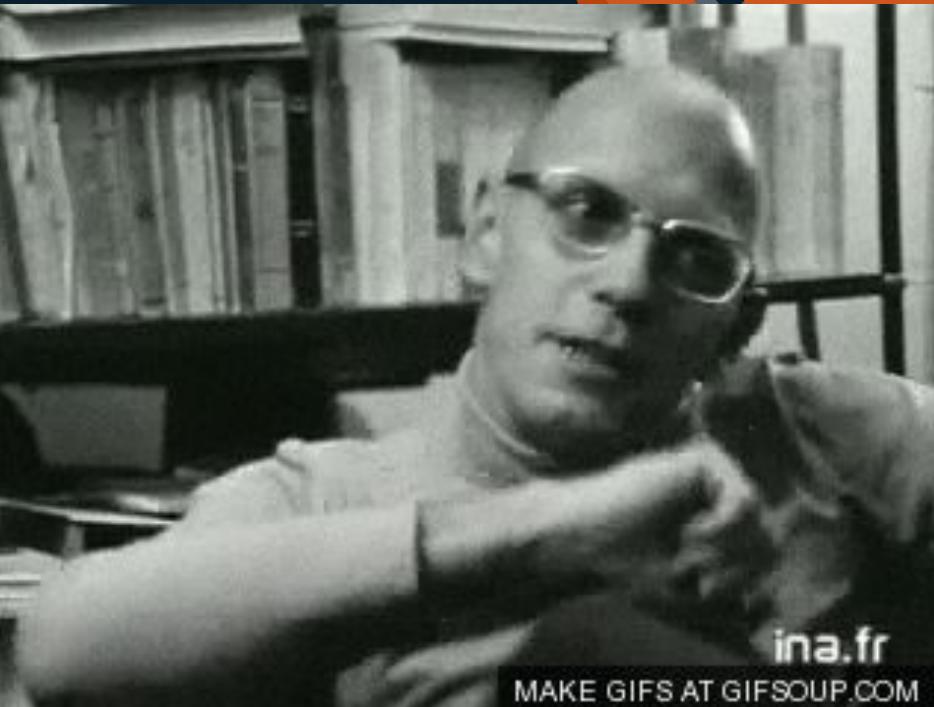


SCHOOL OF
ADVANCED STUDY | UNIVERSITY
OF LONDON



What I do currently?

I studied History. A lot
of Foucault and no math
or coding.



ina.fr

MAKE GIFS AT GIFFSOUP.COM

How I ended up a historian among the machines?



Large DH projects portfolio



PoliticalMashup



Intermezzo: How to asses DH projects?

Or: Do you want to collaborate with them as RSE?

- **Research or Infrastructure** project?
- **Collaboration** between disciplines or transdisciplinary?
- What are the promised **outcomes** of the projects?
Primarily academic? Or does it include other outputs, such as models, tools and data?

Intermezzo: How to asses DH projects?

Or: Do you want to collaborate with them as RSE?

!!! You will be working with **people**, not just data and code. !!!

Gossip is king ;-)

[Home +](#) [Research +](#) [Research projects](#)

Living with machines

A five-year research project that will take a fresh look at the well-known history of the Industrial Revolution using data-driven approaches

Learn more ↓

Related programmes

[Data science for science](#)

[Research Engineering](#)

Why Living with Machines?

- Largest project I've worked on
- Idea of “radical collaboration”
- Reflective and iterative project design

Learn more ↓

Related programmes

[Data science for science](#)

[Research Engineering](#)

Context: Living with Machines

- Collaboration between the Alan Turing Institute and the British Library
- Partner institutions: Cambridge, East Anglia, Exeter, Queen Mary, University of London
- Large, interdisciplinary team comprising data scientists, librarians, historians, and yours truly



Our Partners

Our Funders

The “Living with Machines” Team

Principal and Co-Investigators



Ruth Ahnert
(QMUL)



David Beavan
(Turing)



Emma Griffin
(UEA)

Project team



Claire Austin
Rights Assurance



Kaspar Beelen
Digital Humanities Senior Research Associate



Mariona Coll Ardanuy
Computational Linguistics Senior Research Associate



André Piza
Research Project Manager



Griffith Rees
Research Data Scientist



Kalle Westerling
Research Software Engineer



Timothy Hobson
(Turing)



Jon Lawrence
(Exeter)



Maja Maricevic
(British Library)



Lélie Demertzí
Programme Coordinator



Luke Hare
Research Data Scientist



Sherman Lo
Research Data Scientist



Daniel Wilson
History Senior Research Associate



Rosie Wood
Research Data Scientist



Barbara McGillivray
(Turing / King's College London)



Mia Ridge
(British Library)



Alan Wilson
(Turing)



Katie McDonough
Senior Research Associate



Federico Nanni
Senior Research Data Scientist



Nilo Pedrazzini
Corpus-Based Digital Humanities Research Assistant

Our Partners



The
Alan Turing
Institute

UNIVERSITY OF
CAMBRIDGE

UEA University of
East Anglia

UNIVERSITY OF
EXETER

Queen Mary
University of London

Our Funders

Arts & Humanities
Research Council

UK Research
and Innovation

Living with Machines is...

Past machines: Explores the **social and cultural impact of mechanisation** by investigating digitized historical collections

- Newspapers, maps and census returns
- Examples...





Find
my past

in association with
THE NATIONAL ARCHIVES

1921 CENSUS
UNLIMITED ACCESS

Find out more



History's colourful stories in black and white

Explore millions of digitised newspaper pages, brought to you by Findmypast and the British Library

Search people, places, events and more



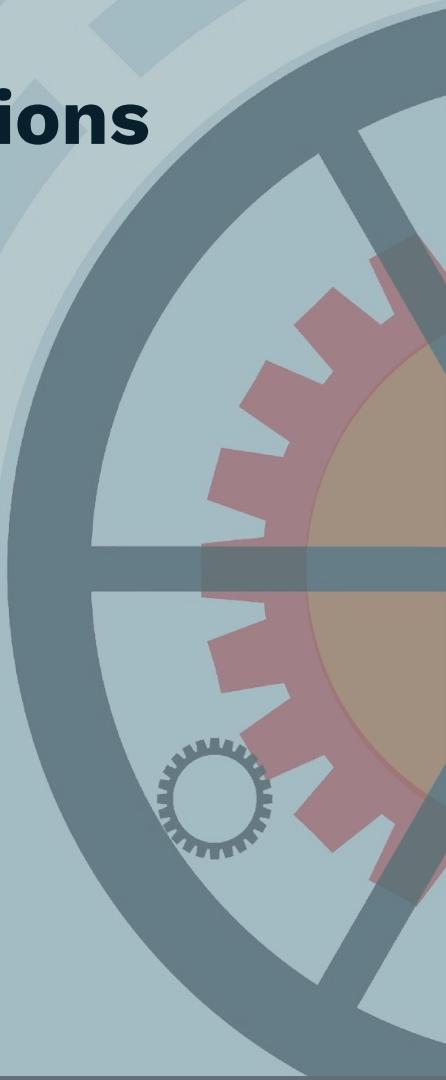
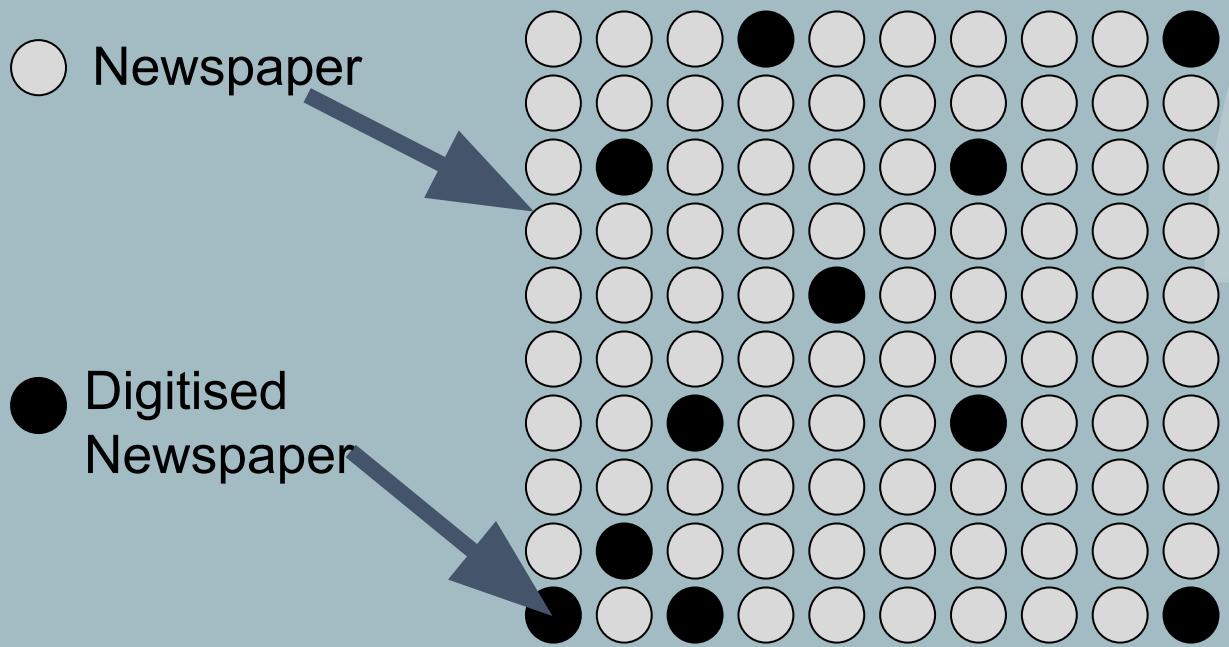
Advanced Search

Browse by Date Country Region County Place Recently Added titles

67,771,346 Pages

Dating from the 1700s

Bias in digitized newspaper collections



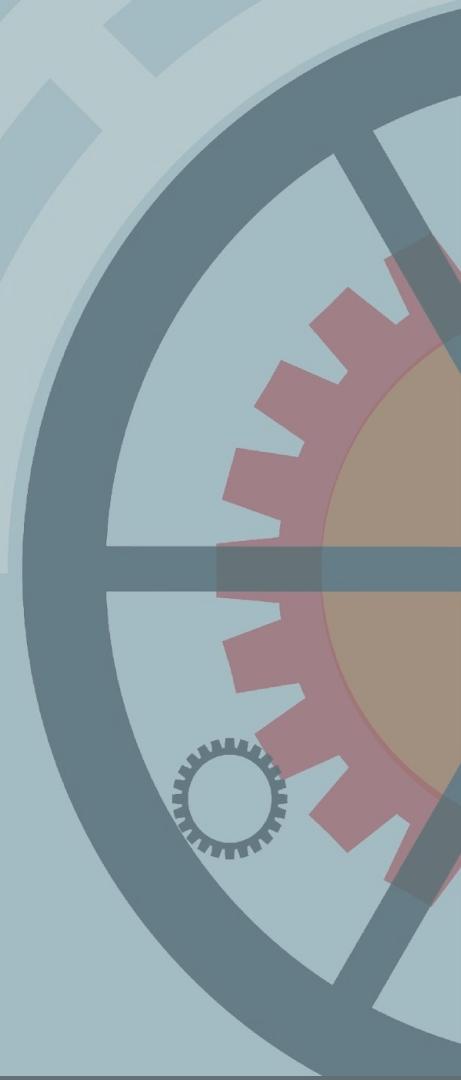


Over-represented



Under-represented

Partial KL



Research on maps and ‘railspace’



1888

How do researchers work with maps?



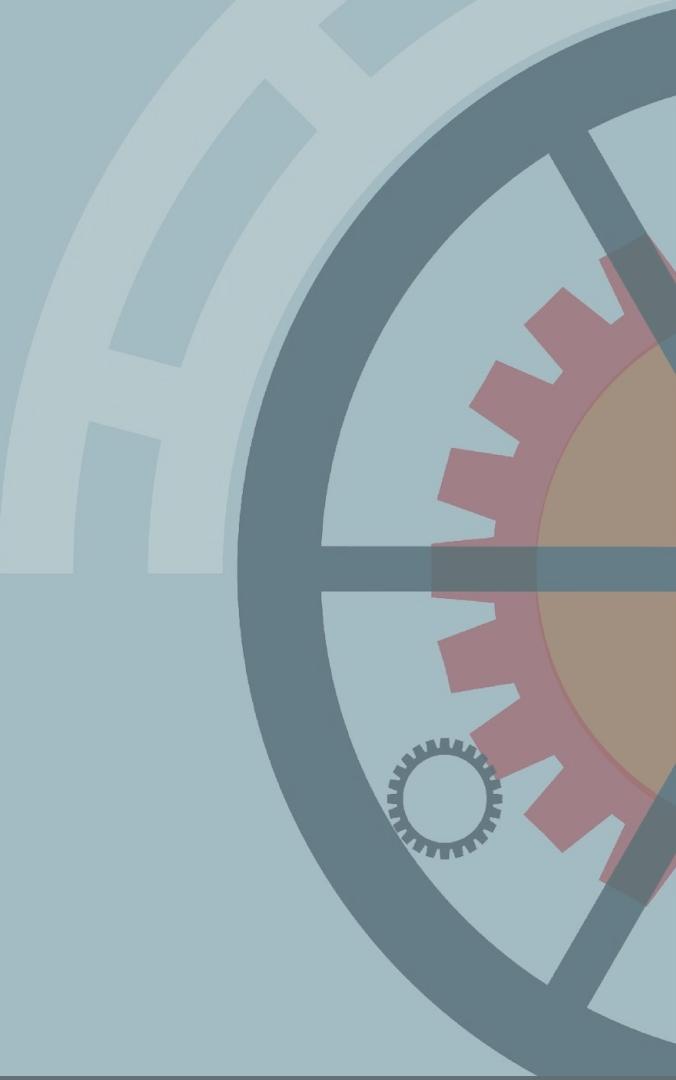
Digitized maps can be more than sheets to browse in a virtual reading room. *But how?*

Ordnance Survey
maps of England,
Wales, and
Scotland

6 inches to 1 mile
1888-1913
(2nd edition)

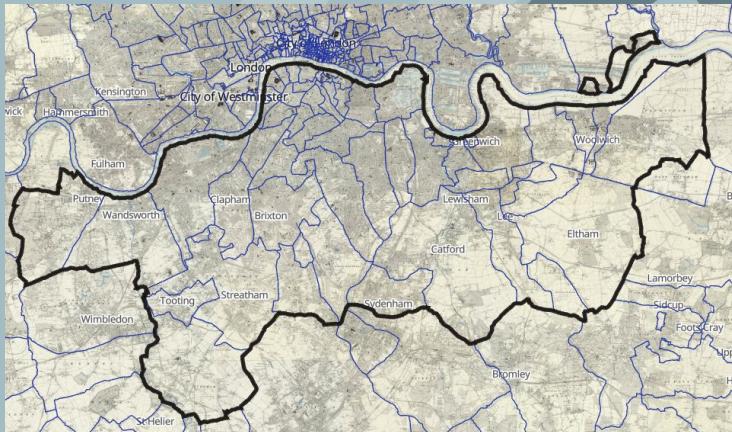
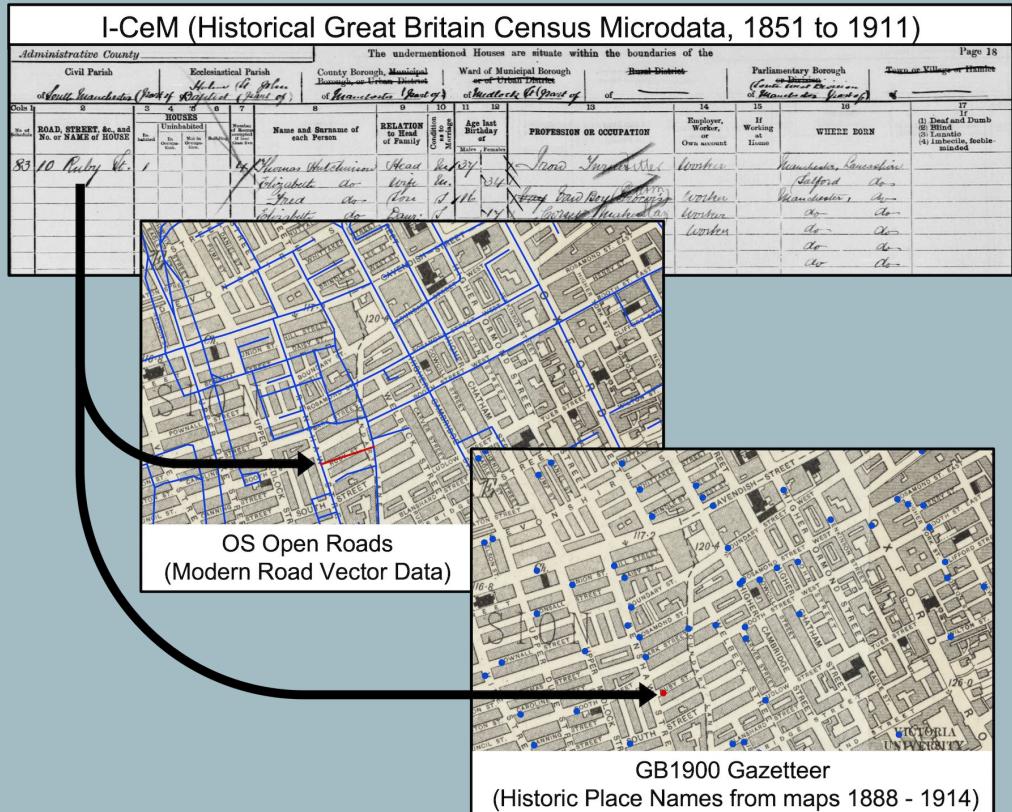
~15K sheets

Linking census records



Geo-coded individual historic census data

(currently England & Wales 1901 - adding Scotland)



Streets x railspace

Railspace

Streets as a proportion of railspace patches within 200m:
blue is low density
yellow/orange mid density
red high density

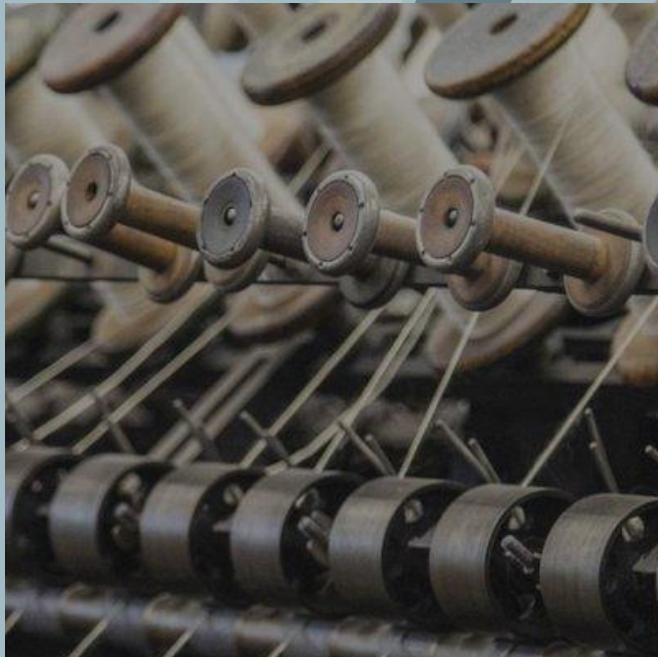


Research on the language of mechanisation

-> Fede?

Living with Machines is...

Present machines: An investigation into what it **means** to use computational analysis for history



Living with Machines is...

Present machines: history and the digital

- Changing research **instruments** (methods, code and software)
- Changing **scales** of research and **source criticism**
- Changing **research practices** related authorship, outputs and **collaboration**



Collaboration within the Living with Machines project

Main resource was my ‘goodbye email’ to the LwM team.
Almost all points I mentioned related to how we collaborated.

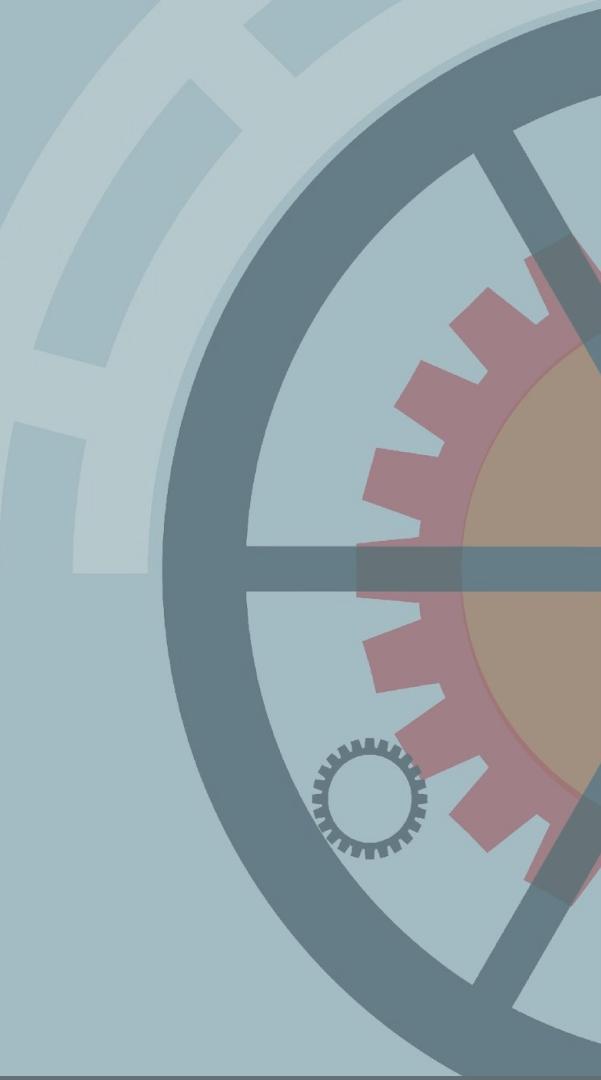
[...] I want to share some important insights and lessons I have learned over the last four years (in random order, more or less):

- Appreciate the hard work/support that is not always immediately visible
- Cherish those months in which the last Friday falls on the 26th
- Copyright is real
- It's OK to say "no". But you should still feel guilty about it (a bit...)
- Good management is a blessing
- Green bananas are NEVER tasty (you might as well chew on a potato)
- In case of doubt, go for ❤️ team clementine ❤️
- Not all good ideas become a reality
- Not all realities were necessarily a good idea
- The distance between the nearest Wasabi and the BL staff entrance is exactly 549 steps
- Changing your mind about things is a great skill (...maybe)

[continued] Unresolved issues and things I am still grappling with:

- Will this email address still work in one hour?
- **Unicodede decode errors (classic one, will never happen)**
- **How to annotate**
- Where did the delicious, rectangular shortbread cookies go?
- Twitter, really... ?
- **What do machines actually learn?**
- Setting the timer for the second floor microwave
- **I wish I were two dogs, then I could play together. ([Godfried Bomans](#))**

[...]



Collaboration within LwM

What does ‘radical collaboration’ mean

“[...] a form of collaboration in which no discipline or professional practice was **in service** of another, and one in which we empowered all researchers to bring their expertise to the table.” (Ahnert et al., 2023, p. 64)

- Issue: How to marry the **‘waterfall’** model of the project with **agile/bottom-up** ways of collaborating
- Developing **‘trading zones’** and **‘contact languages’** (Galison, Kemman): cross-disciplinary communication requirement for effective collaboration
- LwM **reflective** and **iterative** projects: **how we work together important part of the research and project**

Collaboration within LwM: Getting Started

Project Charter

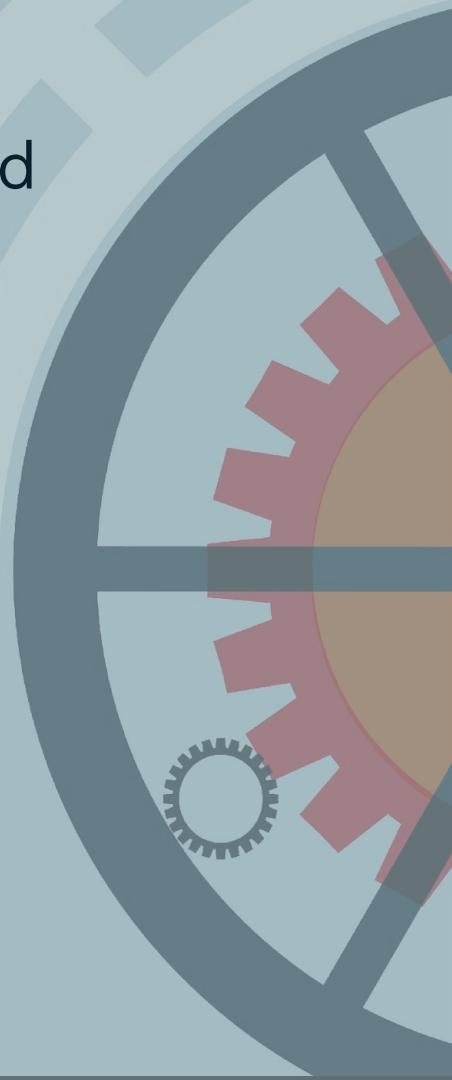
- Outline the project culture and ethos: “Collaboration can be **uncomfortable** [...] it takes a proactive approach, which can be time-consuming [...]"
- The charter paves the way for a ‘**contact language**’ by listing core-values for the team
- Based on Ruecker, S., and Radzikowska, M. (2008)*, revised during the project

Ruecker, S., and Radzikowska, M. (2008). ‘The Iterative Design of a Project Charter for Interdisciplinary Research’. 288–94.
<https://dl.acm.org/doi/10.1145/1394445.1394476>.

Collaboration within LwM: Getting Started

Project Charter

“We are interested in disseminating the results of this project as **widely** and **openly** as possible, with **credit** to **us** for doing so. Our policies around credit should **balance both generosity and meaningfulness.**”



Collaboration within LwM

Credit and Authorship

- Practices **differ** across disciplines
- When you work as part of a large team, where to draw **boundaries** between **authors**, **contributors**, and **consultants**
- Traditionally an emphasis on **writing**, but what about **'hidden'** and **'enabling work'**: “the broader hinterland of labour that goes into the production of data-driven and computational research” (Ahnert et al., 2023, p. 67)
- No '**automatic**' authorship



Collaboration within LwM

CRediT (Contributor Role Taxonomy)

Living Machines: A study of *atypical animacy*

Conceptualization

Mariona Coll Ardanuy^{1,5}
Daniel CS Wilson^{1,5}

Methodology

Mariona Coll Ardanuy
Federico Nanni¹
Kasra Hosseini¹

Implementation

Federico Nanni
Kasra Hosseini
Mariona Coll Ardanuy
Kaspar Beelen^{1,5}

Reproducibility

Kasra Hosseini
Federico Nanni
Mariona Coll Ardanuy
Kaspar Beelen

Interpretation

Kaspar Beelen
Mariona Coll Ardanuy
Katherine McDonough^{1,5}
Daniel CS Wilson
Ruth Ahnert⁵
Jon Lawrence⁴
Giorgia Tolfo²

Historical Analysis

Daniel CS Wilson
Katherine McDonough
Kaspar Beelen
Jon Lawrence

Data Curation

Kaspar Beelen
Mariona Coll Ardanuy
Federico Nanni
Giorgia Tolfo

Annotation

Giorgia Tolfo
Ruth Ahnert
Kaspar Beelen
Mariona Coll Ardanuy
Jon Lawrence
Katherine McDonough
Federico Nanni
Daniel CS Wilson

Writing and Editing

Mariona Coll Ardanuy
Federico Nanni
Ruth Ahnert
Kaspar Beelen
Kasra Hosseini
Jon Lawrence
Katherine McDonough
Barbara McGillivray^{1,3}
Daniel CS Wilson

Supervision

Barbara McGillivray
Ruth Ahnert

Project Management

Barbara McGillivray
Ruth Ahnert

Collaboration within LwM

Credit and Authorship: Values

“We want to credit **all parts of the workflow** in our publications and other outputs through authorship or citation of preceding outputs.”

“All outputs should be regarded as **open to anyone who wants to be involved** but author credit only comes with a **substantive input** [...]”

“Our decisions on authorship questions need to **recognise the venue in which** they are to appear, and the norms of publication in that field.”

Collaboration within LwM

Training, Learning and Researching together

- “We value meeting in person (where at all possible), and meeting regularly, in order to build community, shared understanding, and expertise.”
- Creating physical and/or intellectual “spaces for exchange”:
 - Skill-sharing and peer-training: “Code and Coffee”
 - Collaborative thinking: “Hypothesis Generation”
 - Researching in “Labs” in an agile way

Collaboration within LwM

Minimum Research Outcome

Minimal Viable Product (Eric Ries, 2009): “a product with just enough features to satisfy early customers, to test hypotheses, and to provide feedback for future product development.”

- Time-boxing experimentation and resources
- Outcome vs outputs vs product: focus on proof-of-concept, might this work/be worthwhile?
- Improvements
 - focus on how to keep things minimal
 - “Sunk cost fallacy” and “path dependency”

Collaboration within LwM



Moving online: Pros and cons of being boxed into tiny London flats

- “We lost serendipitous encounters for learning, brainstorming and problem-solving that were more natural in physical spaces.” (Ahnert et al., p. 62)
 - > Re: bananas and cookies?
- Loss in creativity/exploration + deeper rabbit holes
- Zooooooooooooooom meeetings takeaaaaaaaaaaaaaa-hages omg omg (when you work with a large team)

Moving online: Pros and cons of being boxed into tiny London flats

- + Collaborative coding and co-writing. Digital frameworks a format for interaction
- + Private lives become more public, empathy for colleague's constraints and requirements
- + Involvement of geographically 'distant' team members
- + Longer dog walks

Let's now turn to...

The Dark Side of
Collaboration*

* This is mostly an exercise of self-criticism, not of others

Or: lessons I wished I learned earlier.

*'There is only light
at the end of the
tunnel if someone
switched it on.'*

*Demand transparency about
results and end goals.*

Don't assume people (even senior colleagues) know what they are doing, because often they don't.

*... and this is fine, as long as
they are honest about it.*

*If a request seems
unreasonable, it probably is.
Call it out.**

**in a culturally appropriate manner*

*'Not
everything is
interesting'*

*Collaboration is often a
compromise between
disciplines.*

A GOOD COMPROMISE
LEAVES EVERYBODY MAD

© 1993 THE NEW YORK DAILY NEWS BY UNIVERSAL PRESS SYNDICATE
MORT ZACHAROFF



*DH projects are complex, and
yes, it's often complicated, but
not everything can be
interrogated**

**applies to technical as well critical lines of
inquiry*

*... And feel free to avoid,
simplify or ignore some issues
along the way.*

Collaboration can lead to better research, but it is not necessarily the most effective way of working.

*'Be a good
Catholic,
confess...'*

In an interdisciplinary research environment, everyone should have imposter syndrome (otherwise, you are not paying attention).

*Jack of all trades,
master of none? Yup,
that's probably me ;-)*

Ignorance = Bliss?

*Sometimes. It is OK not to know.
Job ads are lies.*

*Confession 1: I don't know how
to use Docker (and don't care,
really)*

*Confession 2: I never
understood Unicode Errors in
Python (and still don't really
care)*

'Being technical' is not always the most important skill to collaborate as a RSE in DH projects... and what does 'being technical' mean, anyway?

Conclusion 1: But what should you be good at as RSE collaborator?

- Translation between disciplines and communication
- Defining, prioritizing and solving problems
- Expectation management
- Leading on the overarching research question
This is a contested point I know, but don't let yourself be confined to 'technical' roles or interventions?

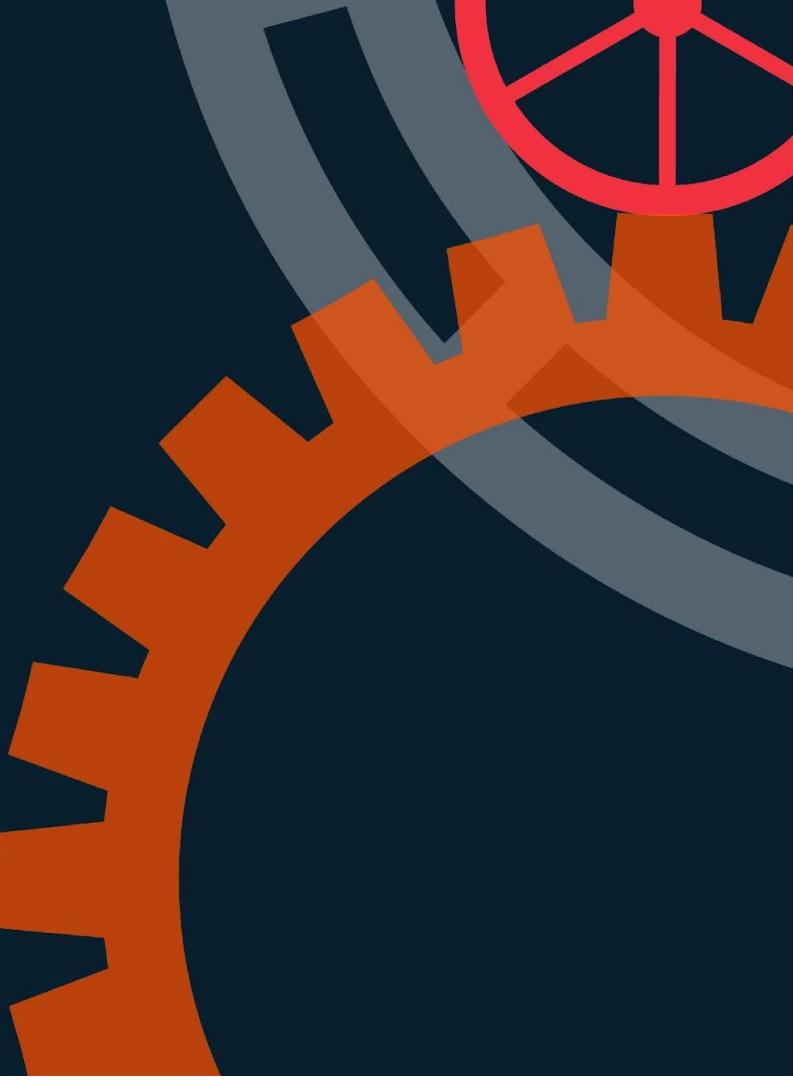
Conclusion 2: Why I still think collaboration is great

- I don't want to grow old alone.
- I have been proven wrong many times, and am grateful for this.
- I gained knowledge of many different fields and topics

Conclusion 2: Why I still think collaboration is great

It's interesting from a purely anthropological
point of view ;-)

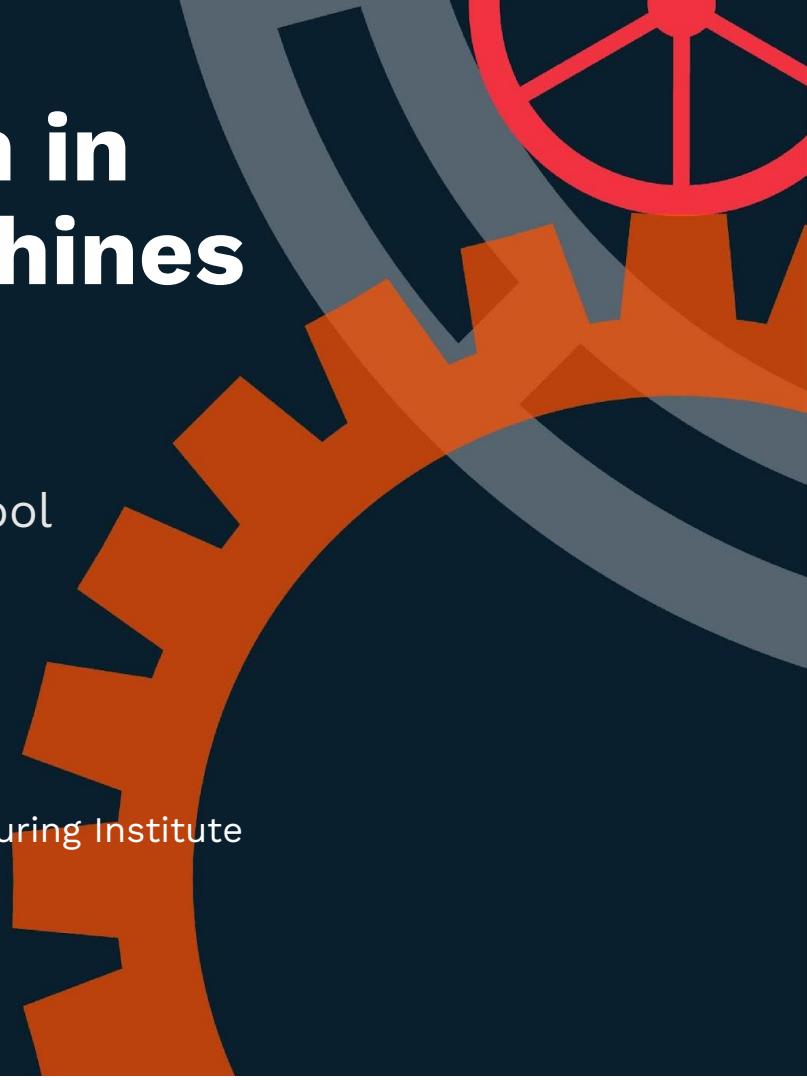
Questions?



NLP Research in Living with Machines

DH & RSE Summer School
5th July, 2024

Federico Nanni
Senior Research Data Scientist, The Alan Turing Institute



Today's talk

1. Brief overview of my **background** and **what I do now**
2. Intro to **The Alan Turing Institute** and **Living with Machines**
3. Overview of a series of works that combine **digital humanities** research and **natural language processing**

Background

BA in **Italian Literature** and an MA in **Contemporary European History**

PhD in **Digital History** at University of Bologna

Topic of my dissertation: considering the web as a historical source

Background

BA in **Italian Literature** and an MA in **Contemporary European History**

PhD in **Digital History** at University of Bologna

Topic of my dissertation: considering the web as a historical source



Dealing with Scarcity of Sources

The screenshot shows the homepage of the Alma Mater Studiorum website. At the top left is the university's circular seal with Latin text "ALMA MATER STUDIORUM" and "A.D. 1088". To the right of the seal is the text "ALMA MATER STUDIORUM A.D. 1088" and "UNIVERSITÀ DI BOLOGNA" in red. On the far right are language links "IT" and "EN". Below the header is a navigation menu with six items: "Home", "Ateneo" (with a subdescription "Storia, strutture, organizzazione"), "Didattica" (with a subdescription "Corsi, iscrizioni, laurea"), "Ricerca" (with a subdescription "Progetti, risultati, trasferimento tecnologico"), "Internazionale" (with a subdescription "Accordi, iniziative, programmi di scambio"), and "Servizi e opportunità" (with a subdescription "Agevolazioni per studenti, orientamento, biblioteche"). A large banner below the menu features a group photo of students at a booth labeled "Associazioni Studentesche". The banner also includes text for "Almaorienta 1 - 2 marzo 2016" and "Vieni a conoscere i nostri corsi, i servizi e le opportunità internazionali". A "continua..." button is at the bottom of the banner. The footer contains a navigation bar with icons for search, user, and cart, followed by page numbers "II" and "1 2 3 4 5".

Nanni, F. (2017). Reconstructing a website's lost past Methodological issues concerning the history of Unibo. *it. DHQ: Digital Humanities Quarterly*, 11(2).

Dealing with Scarcity of Sources



Nanni, F. (2017). Reconstructing a website's lost past Methodological issues concerning the history of Unibo. *it. DHQ: Digital Humanities Quarterly*, 11(2).

Dealing with Abundance of Sources



Explore more than 477 billion [web pages](#) saved over time

<http://>

BROWSE HISTORY

Dealing with Abundance of Sources



Explore more than 477 billion [web pages](#) saved over time

<http://>

BROWSE HISTORY

Nanni, F., Küpper, H., & Ponzetto, S. P. (2016). Semi-supervised textual analysis and historical research helping each other: Some thoughts and observations. *International Journal of Humanities and Arts Computing*.

Nanni, F., Ben-David, A., Brügger, N., Dougherty, M., Milligan, I., & Winters, J. (2016). Web Historiography-A New Challenge for Digital Humanities? In *Proceedings of DH2016*.

Bonfiglioli, R., & Nanni, F. (2015). From close to distant and back: how to read with the help of machines. In *International Conference on the History and Philosophy of Computing*.

From Humanities to Data Science

From Humanities to Data Science



NLP Intern
(Summer 2014)

From Humanities to Data Science



NLP Intern (Summer 2014)

- Basic of **NLP & Python**
- **Evaluation** of methods
- Comparison with **baselines**
- DS language & culture

From Humanities to Data Science



NLP Intern
(Summer 2014)



DATA AND WEB SCIENCE GROUP

Visiting PhD & PostDoc
(Oct 2014 - Oct 2019)

- Basic of **NLP & Python**
- **Evaluation** of methods
- Comparison with **baselines**
- DS language & culture

From Humanities to Data Science



DATA AND WEB SCIENCE GROUP

NLP Intern (Summer 2014)

Visiting PhD & PostDoc (Oct 2014 - Oct 2019)

- Basic of **NLP & Python**
- **Evaluation** of methods
- Comparison with **baselines**
- DS language & culture

- Topic modeling, word embs, DL
- **Publishing** in data science
- **Collaborating** across disciplines
- **Teaching/supervising** students

From Humanities to Data Science



NLP Intern
(Summer 2014)

- Basic of **NLP & Python**
- **Evaluation** of methods
- Comparison with **baselines**
- DS language & culture



DATA AND WEB SCIENCE GROUP

Visiting PhD & PostDoc
(Oct 2014 - Oct 2019)

- Topic modeling, word embs, DL
- **Publishing** in data science
- **Collaborating** across disciplines
- **Teaching/supervising** students



From Humanities to Data Science



NLP Intern (Summer 2014)

- Basic of **NLP & Python**
- **Evaluation** of methods
- Comparison with **baselines**
- DS language & culture



Visiting PhD & PostDoc (Oct 2014 - Oct 2019)

- Topic modeling, word embs, DL
- **Publishing** in data science
- **Collaborating** across disciplines
- **Teaching/supervising** students

The things I enjoyed the most

- Finding clever ways to use “data science” across disciplines
- Collaborating
- Teaching

The things I enjoyed a bit less

- Writing grants
- Publishing papers
- Promoting my research

From Humanities to Data Science (and Back!)



NLP Intern (Summer 2014)

- Basic of **NLP & Python**
- **Evaluation** of methods
- Comparison with **baselines**
- DS language & culture



Visiting PhD & PostDoc (Oct 2014 - Oct 2019)

- Topic modeling, word embs, DL
- **Publishing** in data science
- **Collaborating** across disciplines
- **Teaching/supervising** students

The Alan Turing Institute



Research Data Scientist (Nov 2019 - Now)

- **Best practices** in software eng.
- **Reproducibility**
- **Ethics** and data science
- Working with **historians** as a **data scientist!**

The Alan Turing Institute and The Research Engineering Group





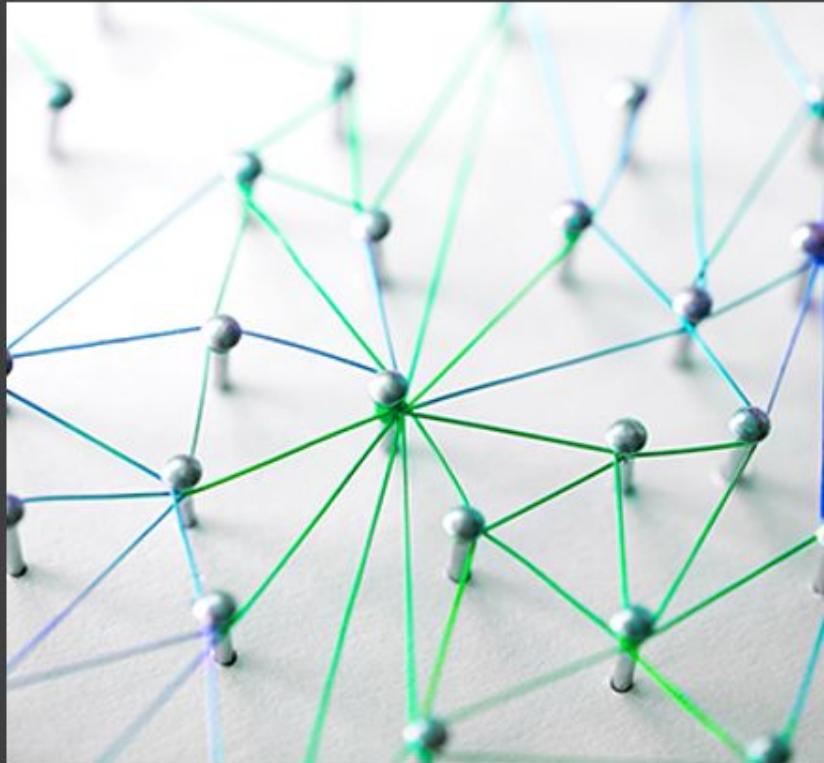
Engineering and Physical Sciences
Research Council

The Alan Turing Institute

Founding the Institute

- “We will found The Alan Turing Institute to ensure Britain leads the way again in the use of big data and algorithm research”
- **George Osborne**, Chancellor of the Exchequer Budget Speech, March 2014

- Network of industry, charity, government partners
- Network of university members
- Strategic government investment

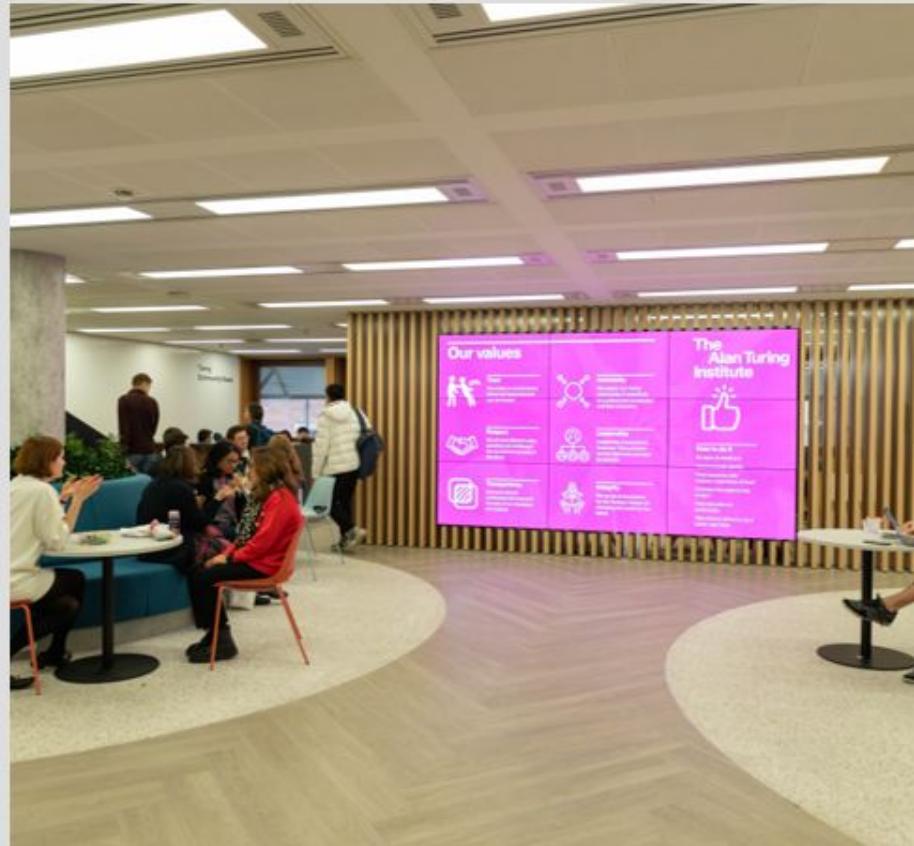




Institute goals

- Advance world-class research and apply it to national and global challenges
- Build skills for the future
- Drive an informed public conversation

-
- 529 staff
 - 58 Enrichment Students
 - 30 Doctoral Students
 - 50 Turing Fellows
 - 27 Turing AI Fellows
 - 11 Turing Research Fellows
 - 10 Honorary Fellows
 - Other external collaborators
(e.g. interest group organisers)



Turing 1.0



Revolutionise healthcare



Deliver safer, smarter engineering



Manage security in an insecure world



Shine a light on our economy



Make algorithmic systems fair, transparent, and ethical



Design computers for the next generation of algorithms

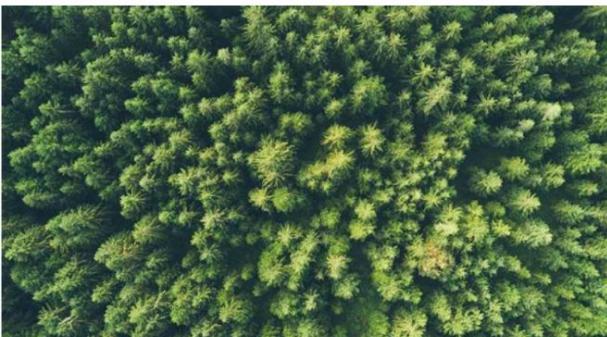


Supercharge research in science and humanities



Foster government innovation

Turing 2.0: Grand Challenges



Defence and national security

To protect the UK, its people and the places they inhabit

Environment and sustainability

To address the climate and biodiversity crisis and the need for greater sustainability.

Transformation of health

To transform health and enable better outcomes for all.

Turing 2.0: Core Capabilities



Fundamental AI



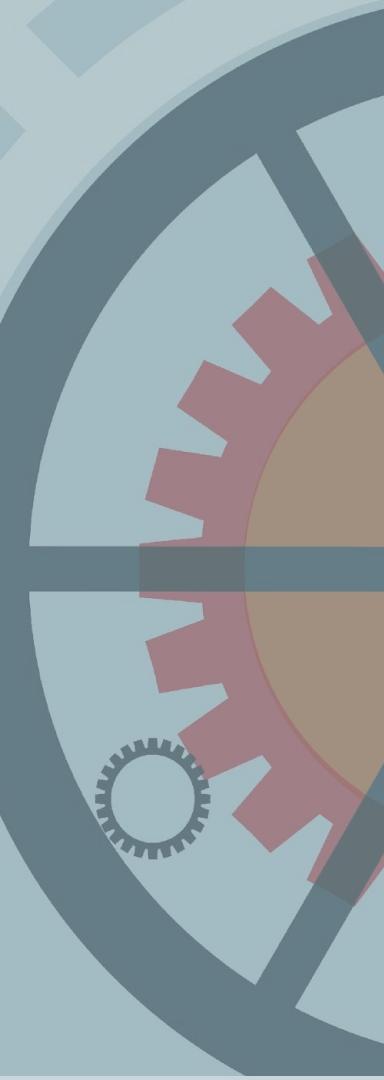
Open-source infrastructure



Research software
engineering capability



Research Application
Managers and Community
Managers



Turing Research Engineering Group

A team of ~45 permanently employed **research software engineers** and **research data scientists**

A **new type of academic role** in the UK, specifically for software developers

Range of backgrounds: physics, biology, computer science, psychology, mathematics, digital humanities...

Enthusiastic **collaborators**, **researchers** and **developers** who want to make long-lasting, reproducible and robust tools and analyses

Remote first team

- Good part of the team outside London.
- Big part hired and onboarded during the peak of the pandemic.
- Half of the team goes to the office regularly
- Collaborators mostly remote
- Focus on documentation, a-synchronicity.



Some of the projects we contribute to



A common interface for discrete choice

Assessing the viability of translating the ALOGIT software to Python and benchmarking it against other discrete choice modelling software



AI for control problems

Using a competition platform to accelerate progress in data-driven control problems



AI for multiple long-term conditions: Research Support Facility

Developing data standards, best practice and community around AI for multiple long term conditions research



AI for Numerical Weather Prediction (NWP): AI4NWP

Developing artificial intelligence algorithms to fundamentally transform weather forecasting and the prediction of local and global weather events



AMBER (AI-assisted Monitoring of Biodiversity using Edge-processing and Remote sensors)

Automated biodiversity monitoring of insects, bats and birds



Analysing humanities books and newspapers data

Text analysis using Cray's supercomputing analytics platform Urika-GX



Analysing social and geographic datasets

Mapping the available tools and techniques for qualitatively understanding human behaviour, in the context of social dynamics



App-based information governance for trustworthy research environments

Delivering an open source 'information governance system in a box' to support data protection in trustworthy research environments



Artificial intelligence for data analytics (AIDA)

Drawing on new advances in AI and machine learning to address data wrangling issues, and help to automate the data analytics process

Living with Machines



Living with Machines

The project ran 2018-2023 and was one of the **biggest** and **most ambitious** humanities and science research initiatives ever to launch in the UK:

- Funded by the AHRC as part of the UKRI Strategic Priorities Fund
- Collaboration between the Alan Turing Institute and the British Library
- Partner institutions: Cambridge, East Anglia, Exeter, Queen Mary

Massive Interdisciplinary Collaboration



Claire Austin
British Library



Dr Kaspar Beelen
Research Associate



Dr Mariona Coll Ardanuy
Research Associate



Dr Karsa Hosseini
Research Data Scientist



Professor Ruth Ahnert
Turing Fellow



David Beavan
Senior Research Software Engineer – Digital Humanities



Professor Emma Griffin
Turing Fellow



Professor Jon Lawrence
University of Exeter



Dr Katherine McDonough
Senior Research Associate



Dr Federico Nanni
Research Data Scientist



André Piza
Research Project Manager, Data Science for Science



Karen Cordier
Research Project Manager (Parental Leave Cover), Living with Machines



Dr Barbara McGillivray
Turing Research Fellow



Maja Maricevic
British Library



Dr Mia Ridge
British Library



Sir Alan Wilson
Director, Special Projects



Dr Giorgia Tolfo
Data and Content Manager, British Library



Dr Olivia Vane
Researcher, British Library



Daniel van Strien
Digital Curator, British Library



Dr Daniel Wilson
Research Associate



Dr Giovanni Colavizza
Visiting Researcher



Dr Adam Farquhar
British Library



Dr James Hetherington
Director of Data Science in Practice



Dr Yann Ryan
British Library



Dr Joshua Rhodes
Research Associate



Dr Sarah Gibson
Research Software Engineer



Dr Rosa Filgueira
Data Architect, EPCC



Dr Timothy Hobson
Senior Research Software Engineer

Massive Interdisciplinary Collaboration



Claire Austin
British Library



Dr Kaspar Beelen
Research Associate



Dr Mariona Coll Ardanuy
Research Associate



Dr Karsa Hosseini
Research Data Scientist



Professor Ruth Ahnert
Turing Fellow



David Beavan
Senior Research Software Engineer – Digital Humanities



Professor Emma Griffin
Turing Fellow



Professor Jon Lawrence
University of Exeter



Dr Katherine McDonough
Senior Research Associate



Dr Federico Nanni
Research Data Scientist



André Piza
Research Project Manager, Data Science for Science



Karen Cordier
Research Project Manager (Parental Leave Cover), Living with Machines



Dr Barbara McGillivray
Turing Research Fellow



Maja Maricevic
British Library



Dr Mia Ridge
British Library



Sir Alan Wilson
Director, Special Projects



Dr Giorgia Tolfo
Data and Content Manager, British Library



Dr Olivia Vanecek
Researcher, British Library



Kalle Westerling
15:54 Today



What the hell!!! I'm not even on this slide!!! lol



Dr Giovanni Colavizza
Visiting Researcher



Dr Adam Farquhar
British Library



Dr James Hetherington
Director of Data Science in Practice



Dr Yann Ryan
British Library



Dr Joshua Rhodes
Research Associate



Dr Sarah Gibson
Research Software Engineer



Dr Rosa Filgueira
Data Architect, EPCC



Dr Timothy Hobson
Senior Research Software Engineer

The Research Engineering team worked across many areas

- Computational infrastructure
- Trusted research environments
- Database
- Data acquisition
- Data analysis
- Machine learning:
 - Maps
 - Text

The Research Engineering team worked across many areas

- Computational infrastructure
- Trusted research environments
- Database
- Data acquisition
- Data analysis
- Machine learning:
 - Maps
 - **Text**



From:
<https://www.theguardian.com/books/2019/mar/25/british-library-planning-leeds-branch-with-boston-spa-upgrade>

LwM people working with textual data

Historians



Professor Ruth Ahnert
Turing Fellow



Dr Katherine McDonough
Senior Research Associate



Dr Daniel Wilson
Research Associate

Digital humanists & librarians



Dr Kaspar Beelen
Research Associate



Daniel van Strien
Digital Curator, British Library

Computational linguists



Dr Barbara McGillivray
Turing Research Fellow



Dr Mariona Coll Ardanuy
Research Associate

Research Engineers



Dr Giovanni Colavizza
Visiting Researcher

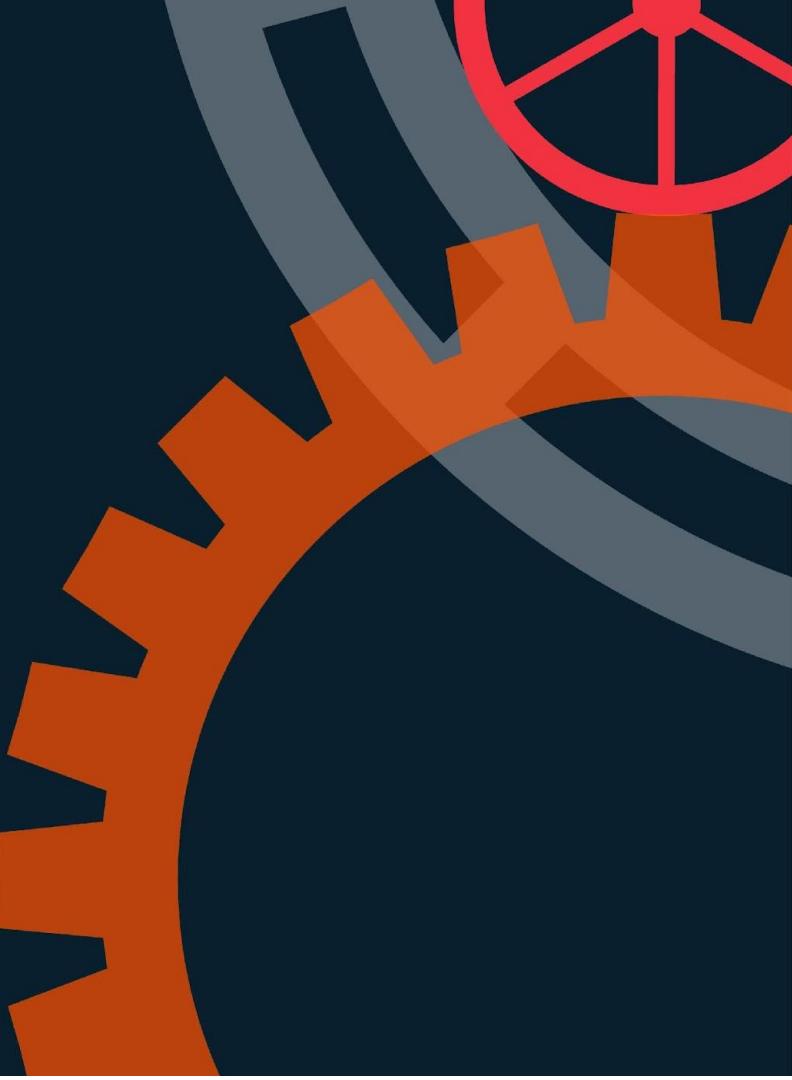


Dr Federico Nanni
Research Data Scientist

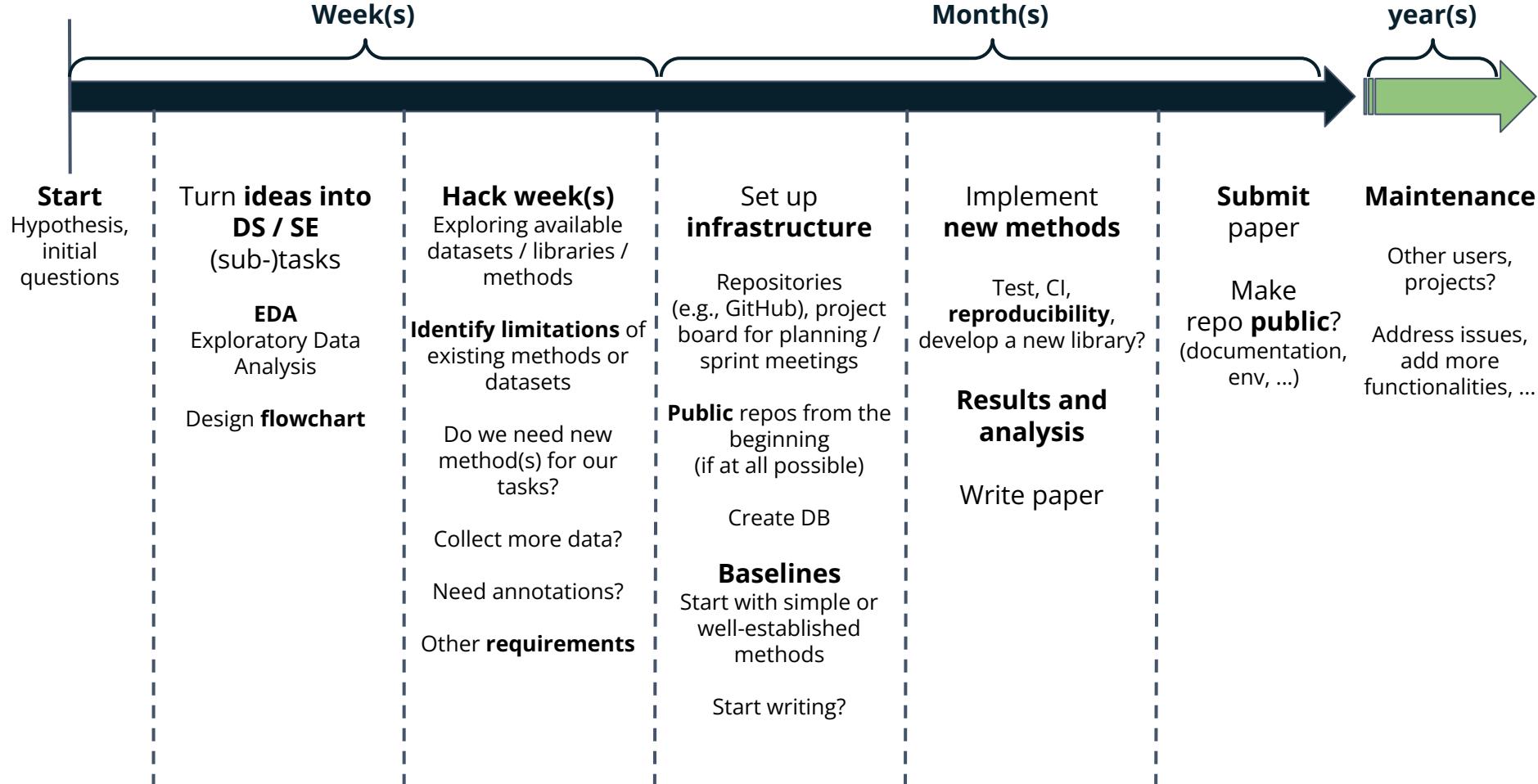


Dr Kasra Hosseini
Research Data Scientist

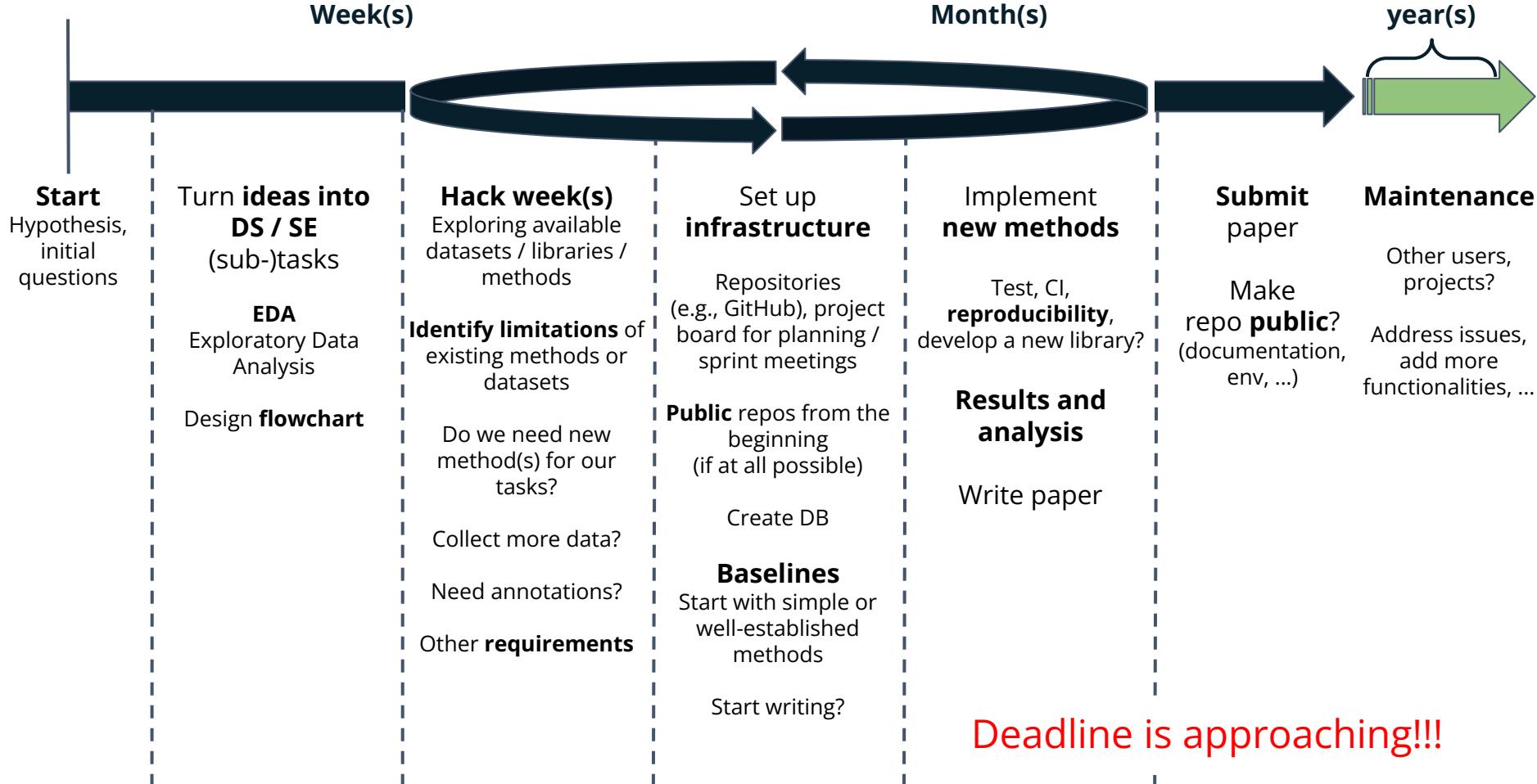
How we worked together (in theory!)



Example timeline



Example timeline



How we worked together (in practice!)



Living Machines

A Study of *Atypical* Animacy
(COLING 2020)

Mariona Coll Ardanuy, **Federico Nanni**, Kaspar Beelen, Kasra Hosseini,
Ruth Ahnert, Jon Lawrence, Katherine McDonough, Giorgia Tolfo,
Daniel Wilson, Barbara McGillivray

Living with Machines project
[@livingwmachines](https://twitter.com/livingwmachines)

Animacy in Linguistics

- Animacy is the **property of being alive**
- **Linguistic animacy** of a given entity tends to align with its biological animacy
 - ... *but not always:*

“He exclaimed; the machine has heard you: it moves!”

The Penny Library of Famous Books, 1895, Publ: George Newnes

- **Machines** sit at the fuzzy **boundary** between animacy and inanimacy (Yamamoto, 1999): deliberate or unconscious

Detecting living machines: motivation

- 19thC Britain: a society being transformed by industrialization
- How machines have been imagined in the 19th century from **lifeless mechanical objects** to **living beings**, and even **human-like agents that feel, think, kill, and love**
- Trace this phenomenon at scale: through time, space, ideologies
- Relevant for today's discussion of the impact of technology in our society (Alan Turing, 1950: “**Can machines think?**”)

19thC Machines animacy dataset

Gathering data to annotate

- Goal: create a dataset of animacy of machines
- Original corpus: 19thC BL Books, ≈48,200, ≈4.9B tokens
- We extracted sentences in English containing machine words (*machine, engine, locomotive...*)
- We extracted interesting sentences through pooling using different methods

19thC Machines animacy dataset

Annotation

Annotation was challenging, even for domain experts.

“No, no, to her mother poor Fraulein was not a woman, a heart, a soul; she was just a machine. ”

Into an Unknown World. A novel, 1897, J.S. Winter

Animacy (true/false): true if the machine is represented as having traits or characteristics (maybe implicit) distinctive of biologically animate beings **or** human-specific skills, feelings, or emotion.

Humanness (true/false): true if the machine is represented as sentient and capable of specifically human emotions.

19thC Machines animacy dataset

- 593 sentences: 201 animate/292 inanimate expressions
- Krippendorff's $\alpha=0.74$ on animacy, $\alpha=0.50$ on humanness.
- Rich in atypical animacy.

Target	Sentence	Animacy	Humanness
engine	In December, the first steam fire engine was received, and tried on the shore of Lake Monona, with one thousand feet of hose.	0	0
engine	It was not necessary for Jakie to slow down in order to allow the wild engine to come up with him; she was coming up at every revolution of her wheels.	1	1
locomotive	Nearly a generation had been strangely neglected to grow up un-Americanized, and the private adventurer and the locomotive were the untechnical missionaries to open a way for the common school.	1	1
machine	The worst of it was, the people were surly; not one would get out of our way until the last minute, and many pretended not to see us coming, though the machine , held in by the brake, squeaked a pitiful warning.	1	1
machines	Our servants, like mere machines , move on their mercenary track without feeling.	1	0
machinery	We have everywhere water power to any desirable extent, suitable for propelling all kinds of machinery .	0	0

Approach in a nutshell

Joy and sorrow - life and death,
wrote the little machine.

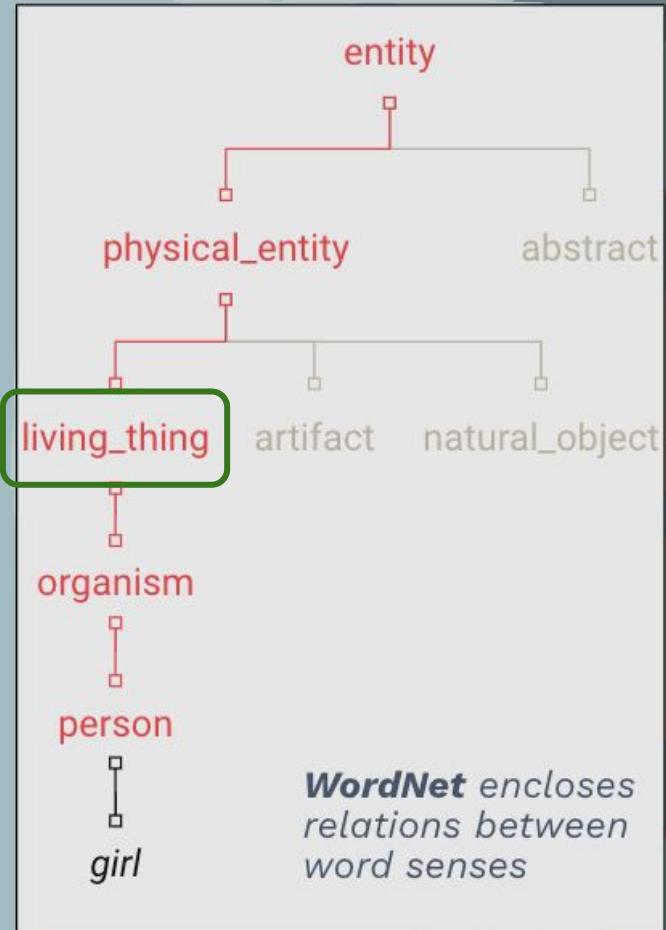
Joy and sorrow - life and death,
wrote the little [MASK].

BERT, predict the missing word in the sentence:

girl	8.1641
man	8.0409
prince	7.4537
one	7.2818
boy	6.9801
princess	6.6766
bird	6.6638
voice	6.6378
lady	6.5472
angel	6.4725
wolf	6.4654
queen	6.3818
witch	6.3068
king	6.2712
sister	6.1635
brother	6.1291

Determining animacy

- Assumption: given a context requiring an animate entity, a contextualized LM predicts tokens corresponding to *conventionally* animate entities.
- For each token in top predicted tokens:
 - Disambiguate to most probable WordNet sense
 - Determine the animacy of the sense using Wordnet hierarchy of nouns
- Threshold and cutoff are found through experimentation.



A Language Model is meant to be a faithful representation of the language that has been used to train it.

“They were told that the [MASK] stopped working.”

BERT language models trained on...

Pre 1850 text:

man 5.3291
prisoners 4.9758
men 4.885
book 4.6477
people 4.556
one 4.4271
slaves 4.4034
air 4.1329
water 4.1148

1850-1875 text:

men 10.7655
people 9.497
miners 9.249
engine 8.0428
women 8.0126
company 7.7261
machine 7.6021
labourers 7.5987
machines 7.5012

1875-1890 text:

men 10.2048
miners 7.6654
machines 7.4062
people 7.2991
engine 7.232
labourers 7.0957
engines 6.7786
engineers 6.5642
machine 6.4712

1890-1900 text:

mercury 8.0446
machinery 7.4067
machine 7.2903
mine 7.274
mill 7.057
men 7.0257
engine 6.9966
lead 6.9177
miners 6.7764

Experiments: baselines

- Most frequent class
- Classification approach
 - Classifiers: SVMs (word embeddings, TFIDF) and BERT Classifier
 - Inputs:
 - **targetExp**: target expression
 - **targetExp + ctxt**: target expression + context (3 token left and right)
 - **maskedExp + ctxt**: masked target expression + context (3 token left and right)
- LSTM sequential tagging approach

Results

	Stories				19thC Machines			
	Precision	Recall	F-Score	Map	Precision	Recall	F-Score	Map
Most frequent class	0.31	0.5	0.383	0.623	0.336	0.5	0.402	0.318
SVM TFIDF: targetExp	0.911	0.893	0.902	0.928	0.696	0.713	0.704	0.474
SVM WordEmb: targetExp	0.927	0.919	0.923	0.954	0.694	0.711	0.702	0.499
BERTClassifier: targetExp	0.951	0.948	0.949	0.985	0.698	0.715	0.706	0.51
SVM TFIDF: targetExp + ctxt	0.734	0.739	0.737	0.859	0.688	0.71	0.699	0.651
SVM WordEmb: targetExp + ctxt	0.758	0.742	0.75	0.876	0.728	0.531	0.614	0.481
BERTClassifier: targetExp + ctxt	0.931	0.926	0.929	0.978	0.695	0.721	0.708	0.721
SVM TFIDF: maskedExp + ctxt	0.674	0.677	0.675	0.804	0.592	0.6	0.596	0.498
SVM WordEmb: maskedExp + ctxt	0.674	0.678	0.676	0.809	0.518	0.52	0.519	0.339
BERTClassifier: maskedExp + ctxt	0.855	0.852	0.854	0.951	0.687	0.696	0.692	0.603
SeqModel: LSTM	0.952	0.948	0.95	0.949	0.697	0.719	0.708	0.482
MaskPredict: BERT-base	0.739	0.703	0.72	0.848	0.719	0.742	0.73	0.74
MaskPredict: BERT-base +ctxt	0.839	0.774	0.806	0.892	0.758	0.778	0.768	0.795
MaskPredict: fit19thBERT +ctxt	—	—	—	—	0.758	0.775	0.766	0.777
MaskPredict: early19thBERT +ctxt	—	—	—	—	0.799	0.773	0.786	0.784

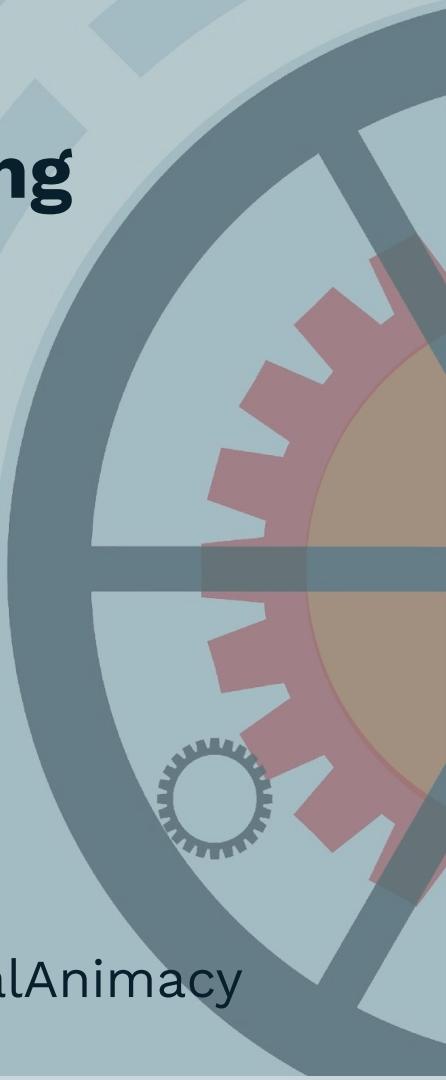
Results

	Stories				19thC Machines			
	Precision	Recall	F-Score	Map	Precision	Recall	F-Score	Map
Most frequent class	0.31	0.5	0.383	0.623	0.336	0.5	0.402	0.318
SVM TFIDF: targetExp	0.911	0.893	0.902	0.928	0.696	0.713	0.704	0.474
SVM WordEmb: targetExp	0.927	0.919	0.923	0.954	0.694	0.711	0.702	0.499
BERTClassifier: targetExp	0.951	0.948	0.949	0.985	0.698	0.715	0.706	0.51
SVM TFIDF: targetExp + ctxt	0.734	0.739	0.737	0.859	0.688	0.71	0.699	0.651
SVM WordEmb: targetExp + ctxt	0.758	0.742	0.75	0.876	0.728	0.531	0.614	0.481
BERTClassifier: targetExp + ctxt	0.931	0.926	0.929	0.978	0.695	0.721	0.708	0.721
SVM TFIDF: maskedExp + ctxt	0.674	0.677	0.675	0.804	0.592	0.6	0.596	0.498
SVM WordEmb: maskedExp + ctxt	0.674	0.678	0.676	0.809	0.518	0.52	0.519	0.339
BERTClassifier: maskedExp + ctxt	0.855	0.852	0.854	0.951	0.687	0.696	0.692	0.603
SeqModel: LSTM	0.952	0.948	0.95	0.949	0.697	0.719	0.708	0.482
MaskPredict: BERT-base	0.739	0.703	0.72	0.848	0.719	0.742	0.73	0.74
MaskPredict: BERT-base +ctxt	0.839	0.774	0.806	0.892	0.758	0.778	0.768	0.795
MaskPredict: fit19thBERT +ctxt	—	—	—	—	0.758	0.775	0.766	0.777
MaskPredict: early19thBERT +ctxt	—	—	—	—	0.799	0.773	0.786	0.784

Results

	Stories				19thC Machines			
	Precision	Recall	F-Score	Map	Precision	Recall	F-Score	Map
Most frequent class	0.31	0.5	0.383	0.623	0.336	0.5	0.402	0.318
SVM TFIDF: targetExp	0.911	0.893	0.902	0.928	0.696	0.713	0.704	0.474
SVM WordEmb: targetExp	0.927	0.919	0.923	0.954	0.694	0.711	0.702	0.499
BERTClassifier: targetExp	0.951	0.948	0.949	0.985	0.698	0.715	0.706	0.51
SVM TFIDF: targetExp + ctxt	0.734	0.739	0.737	0.859	0.688	0.71	0.699	0.651
SVM WordEmb: targetExp + ctxt	0.758	0.742	0.75	0.876	0.728	0.531	0.614	0.481
BERTClassifier: targetExp + ctxt	0.931	0.926	0.929	0.978	0.695	0.721	0.708	0.721
SVM TFIDF: maskedExp + ctxt	0.674	0.677	0.675	0.804	0.592	0.6	0.596	0.498
SVM WordEmb: maskedExp + ctxt	0.674	0.678	0.676	0.809	0.518	0.52	0.519	0.339
BERTClassifier: maskedExp + ctxt	0.855	0.852	0.854	0.951	0.687	0.696	0.692	0.603
SeqModel: LSTM	0.952	0.948	0.95	0.949	0.697	0.719	0.708	0.482
MaskPredict: BERT-base	0.739	0.703	0.72	0.848	0.719	0.742	0.73	0.74
MaskPredict: BERT-base +ctxt	0.839	0.774	0.806	0.892	0.758	0.778	0.768	0.795
MaskPredict: fit19thBERT +ctxt	—	—	—	—	0.758	0.775	0.766	0.777
MaskPredict: early19thBERT +ctxt	—	—	—	—	0.799	0.773	0.786	0.784

A Reproducible Experimental Setting



**Living Machines:
A Study of Atypical Animacy**

License [MIT](#)

This repository provides underlying code and materials for the paper 'Living Machines: A Study of Atypical Animacy' (COLING2020).

Table of contents

- Installation
- Directory structure
- Description of the codes
- Datasets and resources
- Evaluation results
- Citation
- Acknowledgements
- License

<https://github.com/Living-with-machines/AtypicalAnimacy>

The Living Machine: A Computational Approach to the Nineteenth-Century Language of Technology

DANIEL C.S. WILSON, MARIONA COLL
ARDANUY, KASPAR BEELEN, BARBARA
MCGILLIVRAY, AND RUTH AHNERT¹

ABSTRACT: This article examines a long-standing question in the history of technology concerning the trope of the living machine. The authors do this by using a cutting-edge computational method, which they apply to large collections of digitized texts. In particular, they demonstrate the affordances of a neural language model for historical research. In a deliberate maneuver, the authors use a type of model, often portrayed as sentient today, to detect figures of speech in nineteenth-century texts that portrayed machines as self-acting, automatic, or alive. Their masked language model detects unusual or surprising turns of phrase, which could not be discovered using simple keyword search. The authors collect and close read such sentences to explore how figurative language produced a context that conceived humans and machines as interchangeable in complicated ways. They conclude that, used judiciously, language models have the potential to open up new avenues of historical research.

KEYWORDS: history of technology, technology, machines, industrialization, mechanization, labor, automata, language, language models, NLP, text mining, digital humanities, digital history, computational humanities, big data, concept history, nineteenth-century history, methodology



Wilson, Daniel CS, Mariona Coll Ardanuy, Kaspar Beelen, Barbara McGillivray, and Ruth Ahnert. "The Living Machine: A Computational Approach to the Nineteenth-Century Language of Technology." *Technology and Culture* 64, no. 3 (2023): 875-902.

Computational Humanities Research - 2023
Shortlisted for the Best Paper award

The Past is a Foreign Place

Improving Toponym Linking for Historical Newspapers

Mariona Coll Ardanuy, **Federico Nanni**, Kaspar Beelen and Luke Hare
The Alan Turing Institute

A place perspective



A place perspective

Our project has a wide range of **data sources**:

- Time-stamped historical maps,
- Time-stamped historical census,
- Detailed historical data on railway stations, etc!

A place perspective

Our project has a wide range of **data sources**:

- Time-stamped historical maps,
- Time-stamped historical census,
- Detailed historical data on railway stations, etc!

Pinned

 **Living with Machines** @LivingwMachines · 1 Nov 2023

Big news! MapReader has won the 2023 Roy Rosenzweig Prize for Innovation in Digital History, awarded by @AHAhistorians & @chnm. Thank you to the committee for recognizing our work.
[historians.org/awards-and-gra...](https://historians.org/awards-and-grants)
Congrats to all involved!
[@ahrcpress](https://ahrcpress.ac.uk) [@britishlibrary](https://www.bl.uk) [@turinginst](https://turinginst.ac.uk)

Roy Rosenzweig Prize Recipients

The Roy Rosenzweig Prize for Innovation in Digital History is sponsored jointly by the AHA and the Roy Rosenzweig Center for History and New Media (RRCHNM) at George Mason University. This nonresidential prize is awarded annually to honor and support work on an innovative and freely available new media project, and in particular for work that reflects thoughtful, critical, and rigorous engagement with technology and the practice of history.

2022
Katherine McDonough, Daniel CS Wilson, Kaspar Beelen, Kasra Hosseini, Rosie Wood, Andrew Smith, Kalle Westerling, Daniel van Strien, Olivia Vane, Jon Lawrence, and Ruth Ahnert, *MapReader* (*Living with Machines*)

A place perspective

Our project has a wide range of **data sources**:

- Time-stamped historical maps,
- Time-stamped historical census,
- Detailed historical data on railway stations, etc!

Motivation: **linking** the different **datasets by place** enables complex historical analyses across time and space.

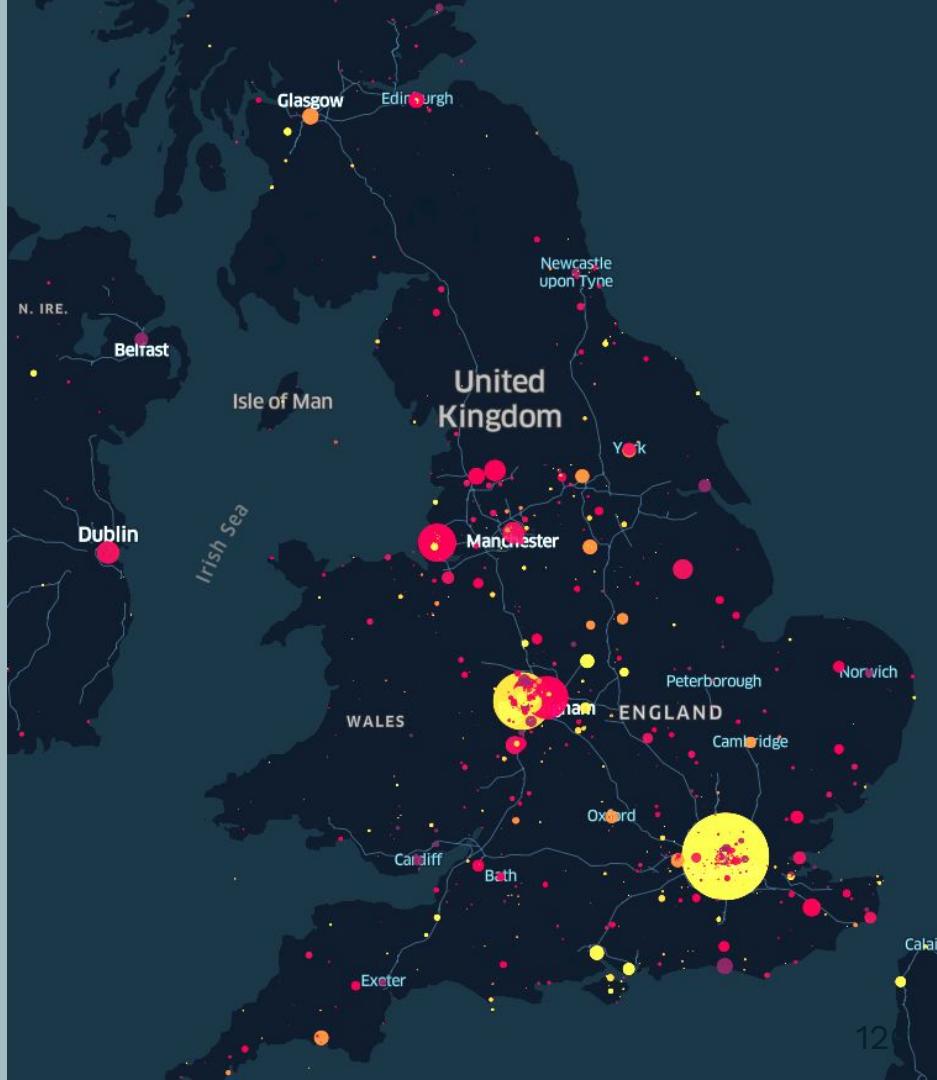
Need: to **geolocate toponyms** (place names) in **newspaper articles**.

Places in newspapers

Places in newspapers

Sample of 4,500 articles
from between 1880 and 1888
from:

*The Stourbridge
Observer, Cradley
Heath, Halesowen &
District Chronicle*



Places in newspapers

Sample of 4,500 articles
from between 1880 and 1888
from:

*The Stourbridge
Observer, Cradley
Heath, Halesowen &
District Chronicle*



Challenges

CxitiTCHUßCit, June 10—Yesterday being the day appointed for the election of two gentlemen to represent this borough in the new imperial Parliament [...] .

Source: *The Dorset County Chronicle* (1864-04-14)

Leghorn, April 6. LETTERS from Constantinople, dated March 3, mention, that an Earthquake had lately happened at Tauris, the Capital of the Province of Syria Ariherbigan, in Persia.

Source: *The Manchester Mercury* (1780-05-30)

Challenges

CxititCHUBCit, June 10—Yesterday being the day appointed for the election of two gentlemen to represent this borough in the new imperial Parliament [...].

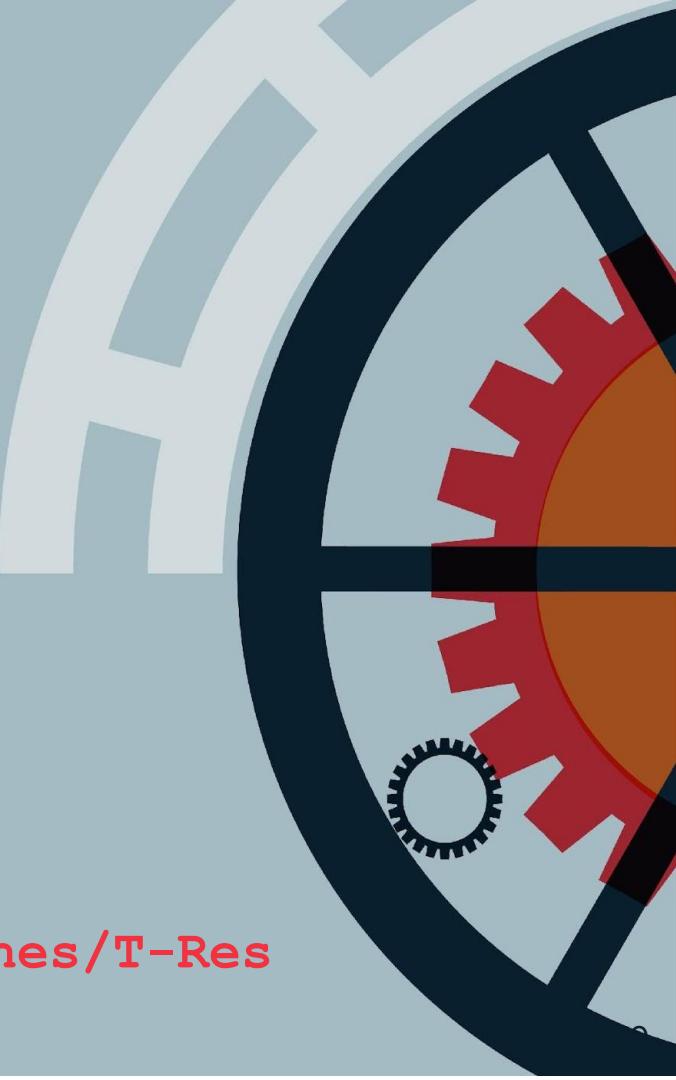
Source: *The Dorset County Chronicle* (1864-04-14)

Leghorn, April 6. LETTERS from Condantinopte, dated March 3, mention, that an Earthquake had lately happened at Tauris, the Capital of the Province of *IS Jra Ariherbigan, in Pcrfia.

Source: *The Manchester Mercury* (1780-05-30)

T-Res: the tool

<https://github.com/Living-with-machines/T-Res>



T-Res

- End-to-end pipeline for toponym resolution.
- T-Res is modular and is built on top of existing robust NLP tools:
 - Transformers for named entity recognition.
 - DeezyMatch for candidate selection and ranking.
 - REL (Radboud Entity Linker) for entity disambiguation.

A geographic knowledge base

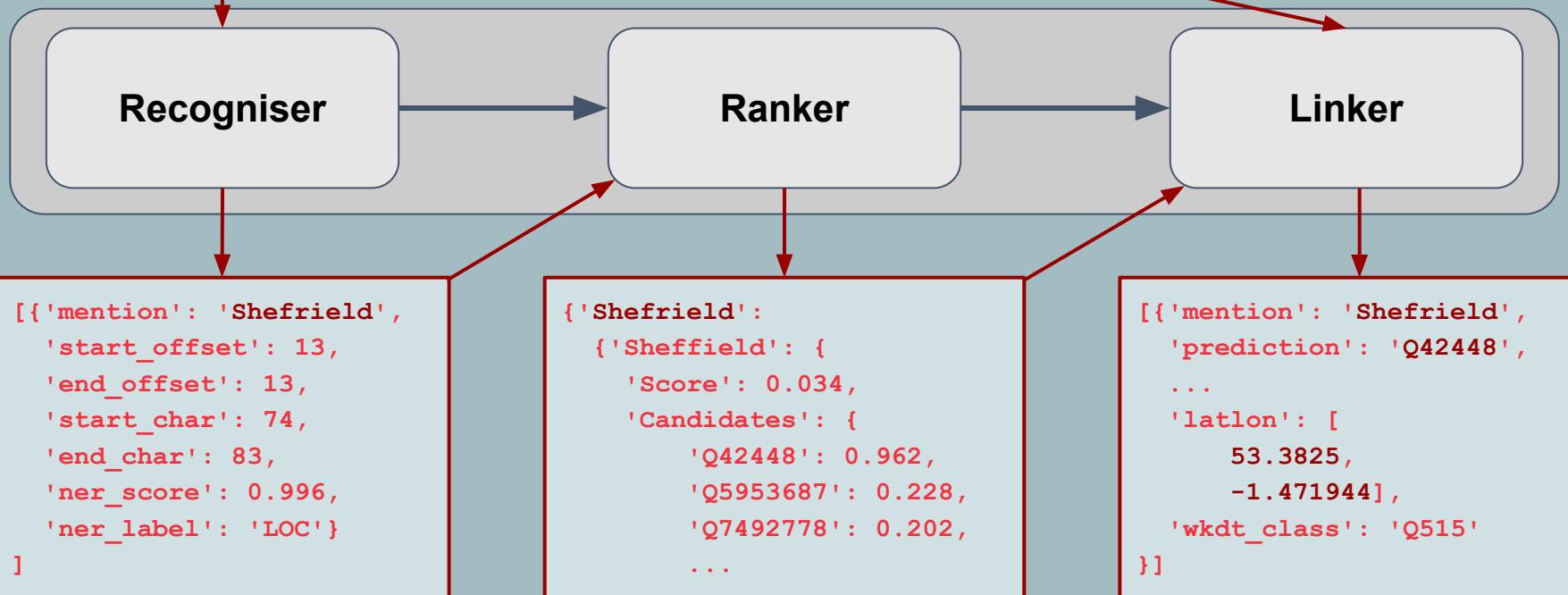
Generated from combining Wikipedia and Wikidata

Wikidata ID	English name	Lat	Lon	Wikipedia mention names	Mention-to-entity frequencies
Q1458	Calabria	39.0	16.5	<i>Calabria region, Region of Calabria, Enotria, Brutium, Calabre, ...</i>	...
Q1467	Cebu City	10.29	123.90
Q1473	Davao City	7.07	125.6
Q1492	Barcelona	41.38	2.18

Source: <https://github.com/Living-with-machines/wiki2gaz>

T-Res Pipeline

"A remarkable case of rattening has just occurred in the building trade of Sheffield."



T-Res Pipeline

"A remarkable case of rattening has just occurred in the building trade of Sheffield."

Recogniser

Ranker

Linker

```
[{"mention": "Sheffield",  
 "start_offset": 13,  
 "end_offset": 13,  
 "start_char": 74,  
 "end_char": 83,  
 "ner_score": 0.996,  
 "ner_label": "LOC"}]
```

```
{"Sheffield":  
 {"Sheffield": {  
 "Score": 0.034,  
 "Candidates": {  
 "Q42448": 0.962,  
 "Q5953687": 0.228,  
 "Q7492778": 0.202,  
 ...}}}}
```

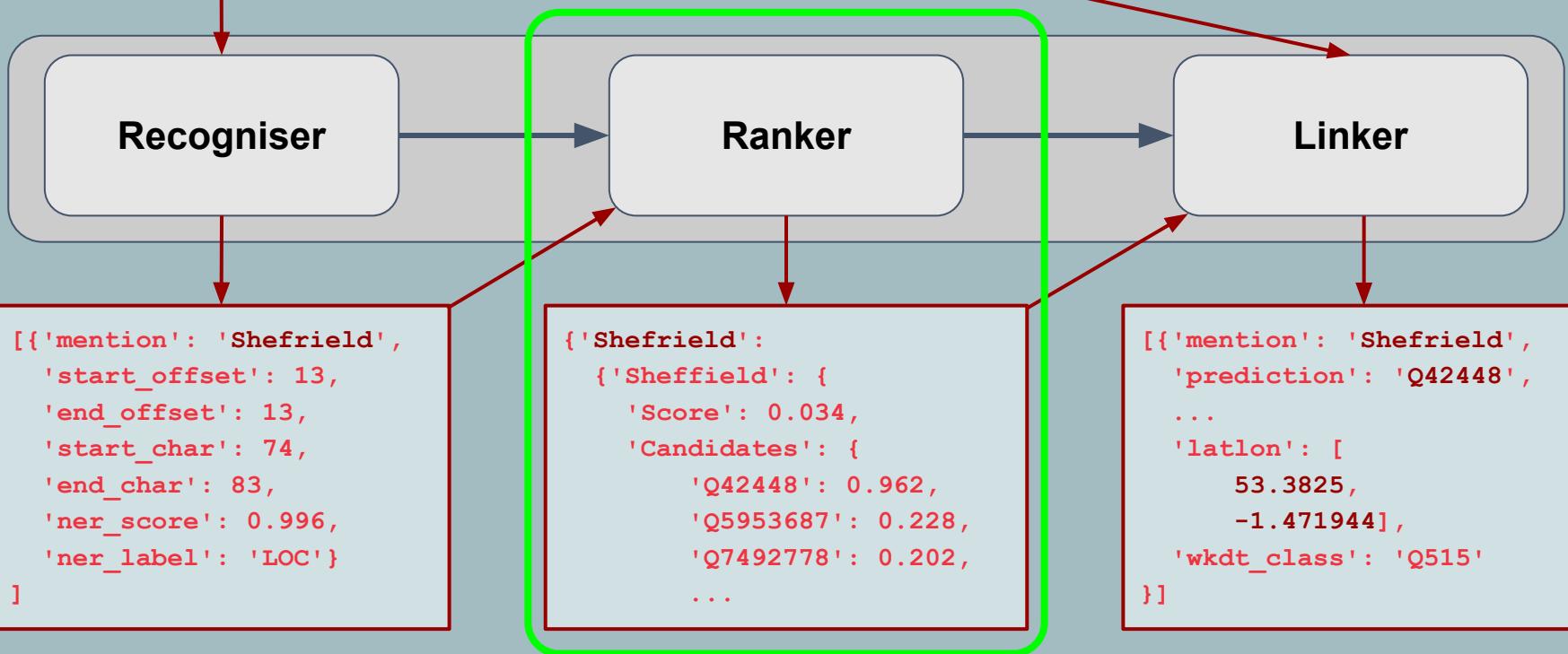
```
[{"mention": "Sheffield",  
 "prediction": "Q42448",  
 ...  
 "latlon": [  
 53.3825,  
 -1.471944],  
 "wkdt_class": "Q515"}]
```

Recognition module

- Built on top of the `transformers` library.
- Users can either:
 - Load an existing BERT-based NER model (locally or from the HF hub).
 - Use T-Res to fine-tune a NER model, given a pre-trained base model and a training set.
- In our experiments:
 - **Base model:** [Livingwithmachines/bert_1760_1900](#) (Hosseini et al 2021)
 - **NER dataset:** [topRes19th](#) training set: BUILDING, STREET, LOC (Coll Ardanuy et al 2022)

T-Res Pipeline

"A remarkable case of rattening has just occurred in the building trade of Sheffield."



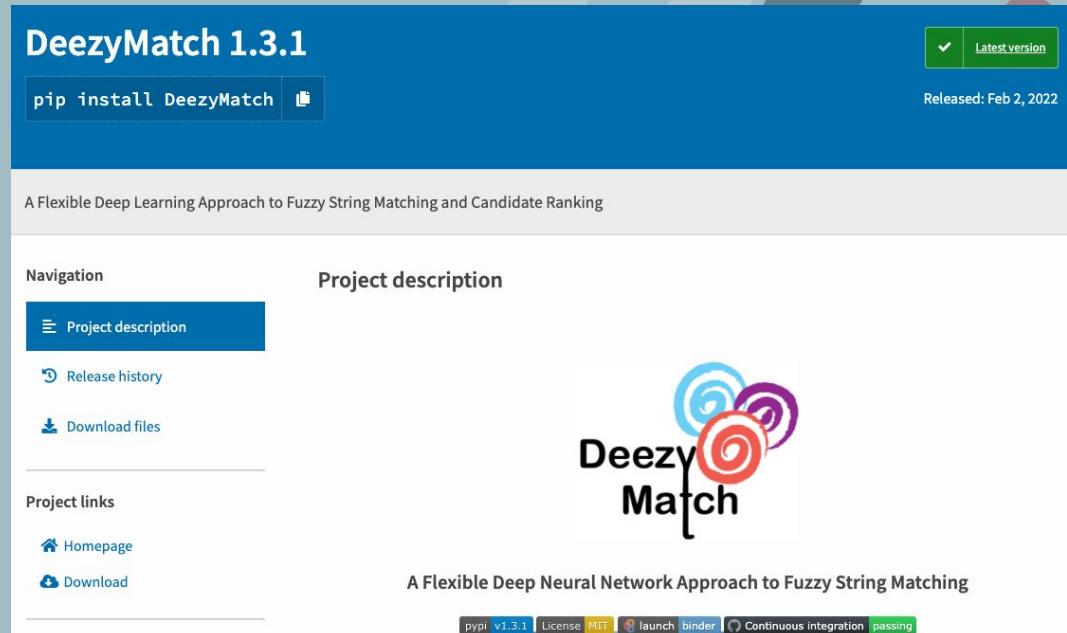
Candidate selection and ranking module

- Strategies for finding candidates in a knowledge base:
 - Exact matching
 - Partial matching based on string overlap
 - Fuzzy string matching based on edit distance (slow!!!)
 - Train and load a DeezyMatch model for fuzzy string matching

Sheffield → Sheffield

DeezyMatch (EMNLP, 2020)

- A **DEE**p learning approach to fuz**ZY** string **MATCH**ing and candidate ranking.
- Python library developed by Living with Machines
- Learns **string transformations** from examples



The screenshot shows the PyPI project page for DeezyMatch 1.3.1. At the top, there's a green button labeled "Latest version". To its right, it says "Released: Feb 2, 2022". Below this, the title "DeezyMatch 1.3.1" is displayed, along with a "pip install DeezyMatch" button. The main content area contains a brief description: "A Flexible Deep Learning Approach to Fuzzy String Matching and Candidate Ranking". On the left, there's a "Navigation" sidebar with "Project description" (which is currently selected), "Release history", and "Download files". On the right, there's a "Project description" section with a logo featuring three colorful swirls (blue, red, purple) and the text "Deezy Match". Below the logo, it says "A Flexible Deep Neural Network Approach to Fuzzy String Matching". At the bottom, there are links for "Homepage" and "Download", along with badges for "pypi v1.3.1", "License MIT", "launch binder", and "Continuous integration passing".

Github repository: <https://github.com/Living-with-machines/DeezyMatch>
Source: Hosseini et al. (2020)

DeezyMatch training examples

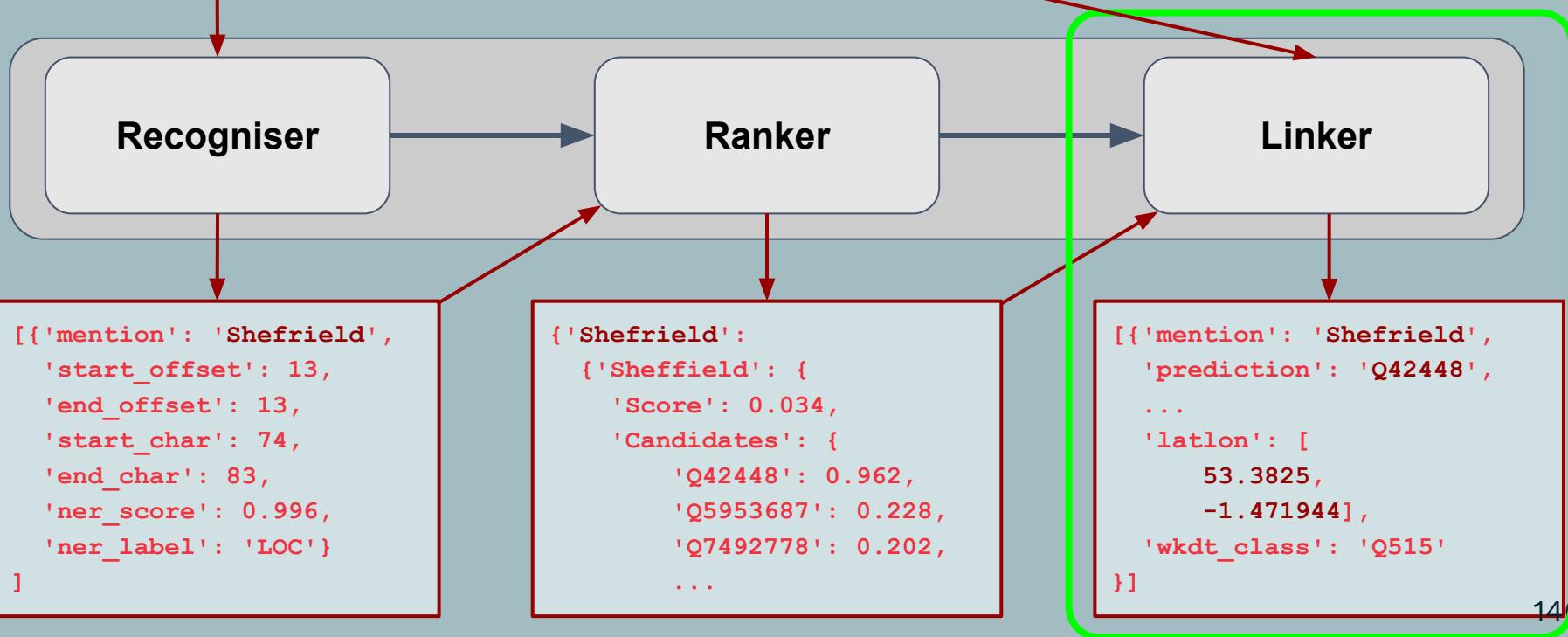
Generated using top-n neighbours
from word embeddings trained on OCR texts

False		True	
would	might	would	would
would	must	would	woull
would	likely	would	wonld
would	thought	would	woubl
would	believed	would	wouid
would	therefore	would	woukl
would	bemuse	would	wouhl
would	certainly	would	worild
...

Source: Pedrazzini & McGillivray (2022): Diachronic word embeddings from 19th-century British newspapers

T-Res Pipeline

"A remarkable case of rattening has just occurred in the building trade of Shefield."



Linking module

Two approaches:

- “**mostpopular**”: selects the candidate that appears as most relevant in Wikipedia (i.e. most frequent entity given a mention).
- “**rel**”: selects the candidate adapting the Radboud Entity Linker (2020) implementation of Le and Titov (2018) disambiguation approach:
 - Option to remove street and building entities.
 - Option to add the place of publication as an already disambiguated entity.

Entity Linking

Selection	Disambiguation	topRes19th			HIPE		
		End-to-end EL			End-to-end EL		
P	R	F1	P	R	F1		
rel-api	rel-api	0.459	0.498	0.478	0.365	0.489	0.418
T-Res:exact	mostpopular	0.552	0.561	0.557	0.377	0.412	0.394
	bydistance	0.170	0.172	0.171	0.085	0.093	0.089
T-Res:deezy	mostpopular	0.588	0.597	0.592	0.462	0.505	0.483
	bydistance	0.177	0.180	0.179	0.126	0.137	0.131
	rel	0.591	0.601	0.596	0.442	0.484	0.462
	rel:nil	0.652	0.663	0.658	0.447	0.489	0.467
	rel+publ	0.579	0.588	0.583	0.452	0.495	0.472
	rel+publ:nil	0.659	0.670	0.664	0.462	0.505	0.483

TABLE OF CONTENTS:

Getting started

[Installing T-Res](#)[Resources and directory structure](#)

The complete tour

[The Pipeline](#)[The Recogniser](#)[The Ranker](#)[The Linker](#)

Reference

[Running T-Res as API](#)[Experiments and evaluation](#)

The complete tour

The T-Res has three main classes: the **Recogniser** class (which performs toponym recognition, or named entity recognition—NER), the **Ranker** class (which performs candidate selection and ranking for the named entities identified by the Recogniser), and the **Linker** class (which selects the most likely candidate from those provided by the Ranker).

An additional class, the **Pipeline**, wraps these three components into one, therefore making it easier for the user to perform end-to-end entity linking.

In the following sections, we provide a complete tour: including an in-depth description of each of the four classes. We recommend that you start with the Pipeline, which wraps the three other classes, and refer to the description of each of the other classes to learn more about them. We also recommend that you first try to run T-Res using the default pipeline, and then change it accordingly to your needs.

The Pipeline

The Pipeline wraps the Recogniser, the Ranker and the Linker into one object, to make it easier to use T-Res for end-to-end entity linking.

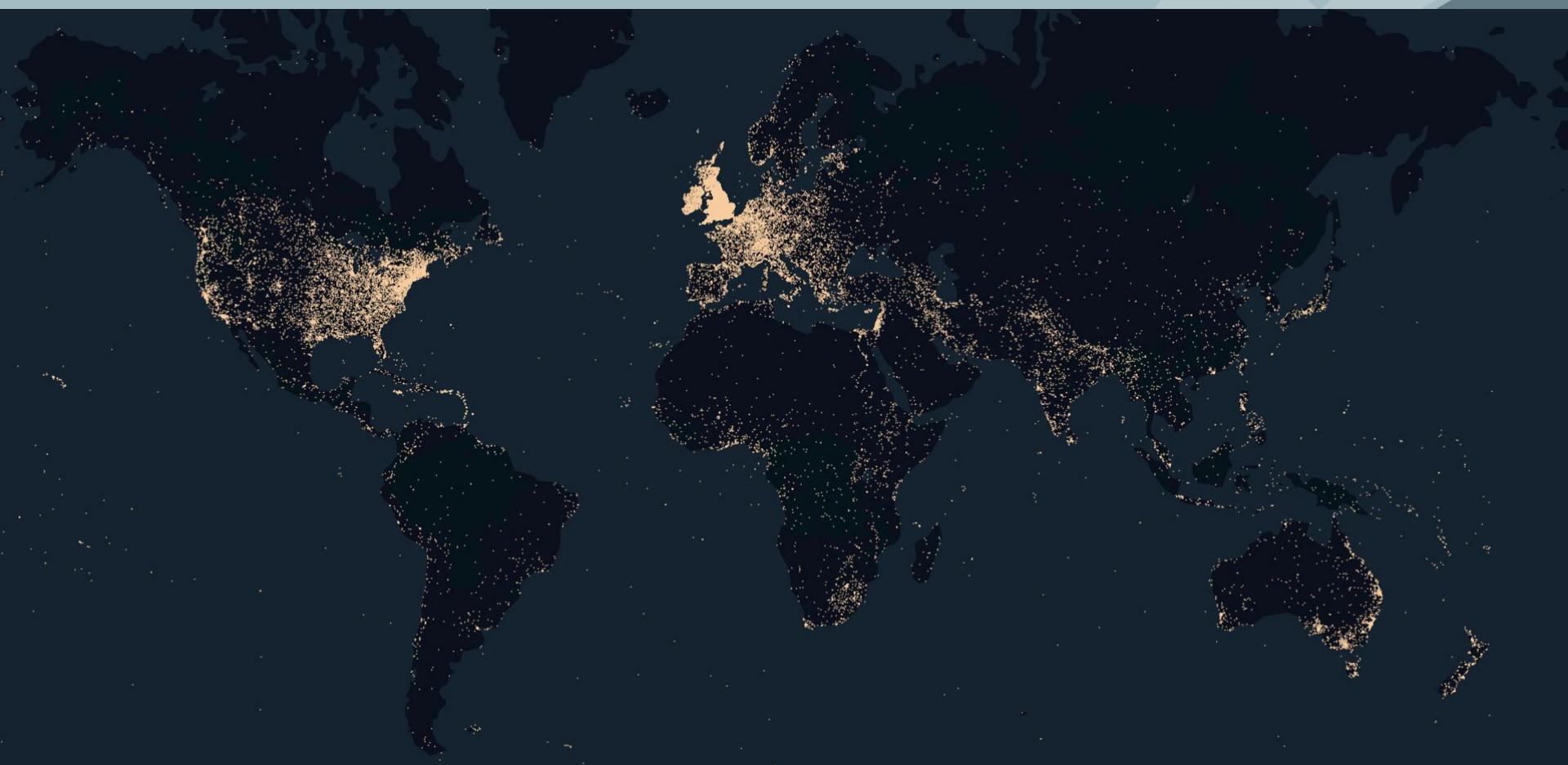
1. Instantiate the Pipeline

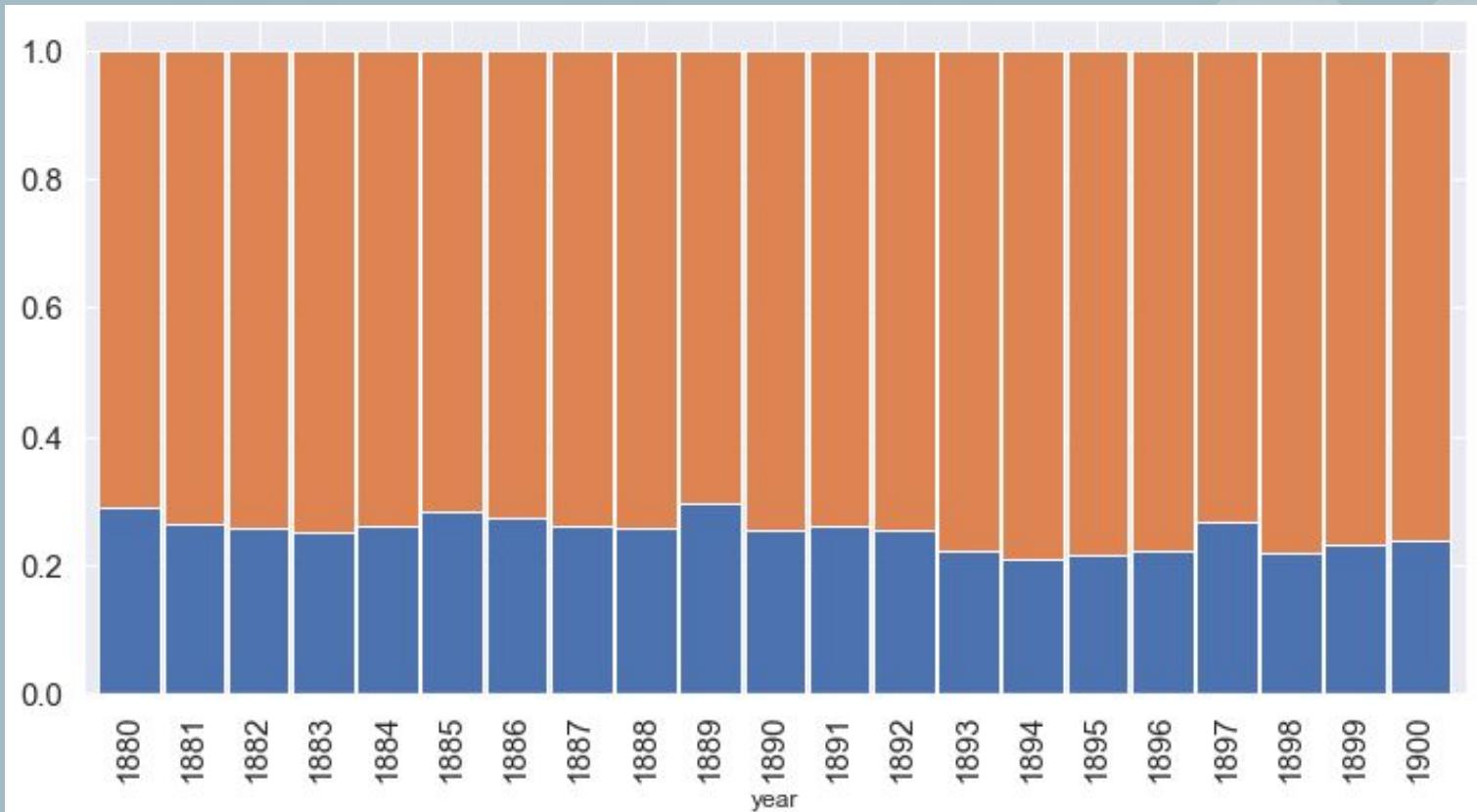
<https://living-with-machines.github.io/T-Res/>

Using the tool in historical research

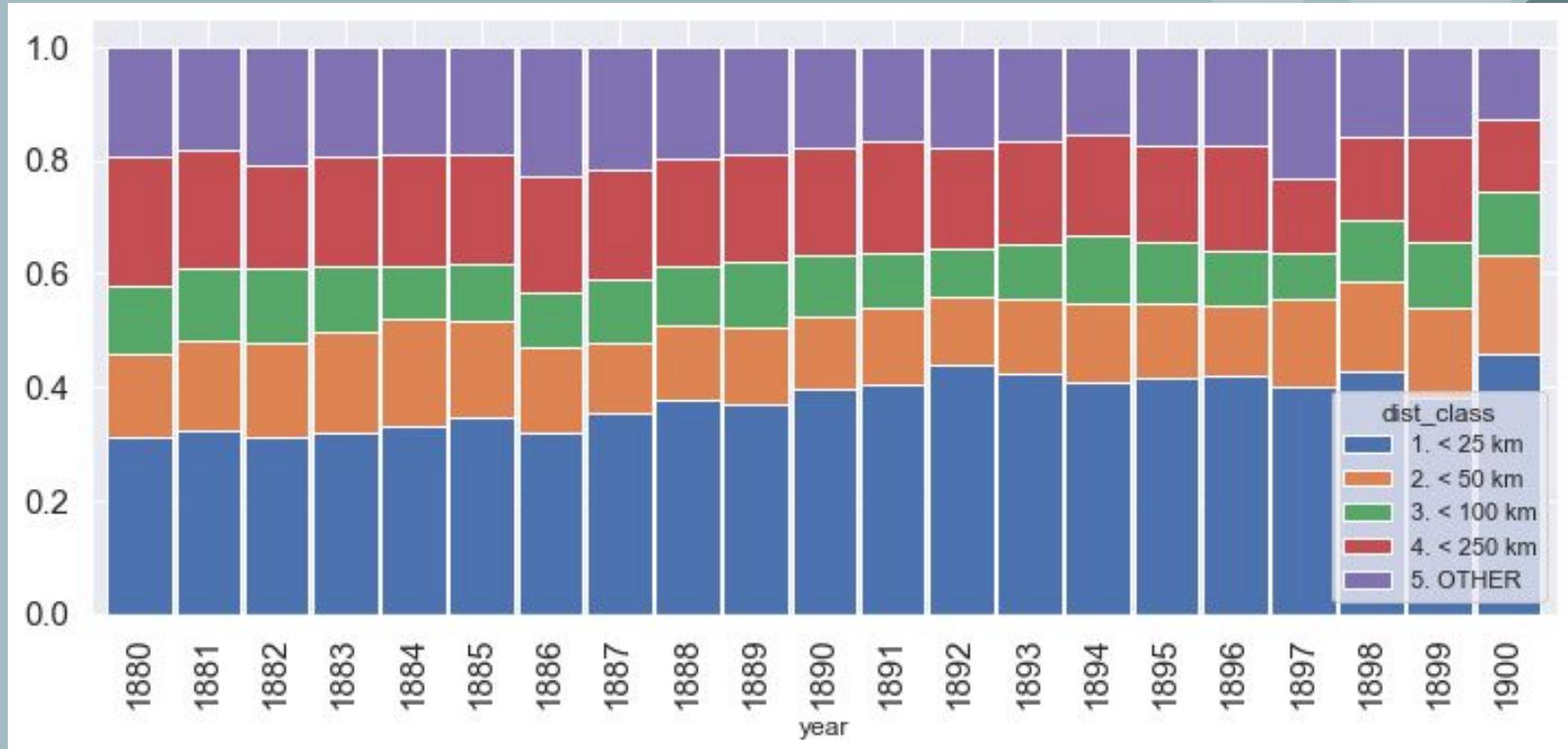
- Preliminary experiments
- Data: Living with Machines
Newspapers
- Different levels from the local
to the transnational







British Toponyms
 Outside the British Isles



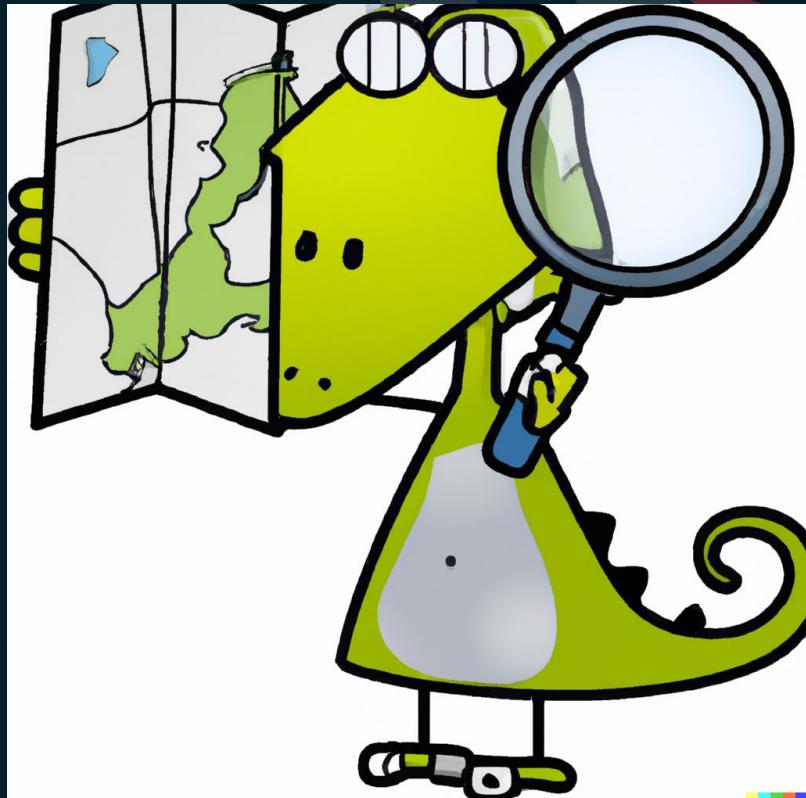
Distance between detected toponym and the newspaper's place of publication

T-Res



<https://github.com/Living-with-machines/T-Res>

Authors: Mariona Coll Ardanuy, Federico Nanni, Kaspar Beelen and Luke Hare, with inputs from the LwM team



Generated with Dall-e: <https://openai.com/research/dall-e>
Prompt: "A cartoon of a funny T-Rex reading a map with a lense"

Wrap-up and lessons learned



Brainstorming ideas together require lots of interdisciplinary efforts

Some of our attempts:

- *HypGen*: hypothesis generation group
- IdeasLab
- NLP reading group
- Computer vision for digital heritage interest group
- Humanities & data science discussion group

Embedding best research software engineering practices is complex

What we tried:

- Offering git-flow overviews
- Being available for informal support (Code & Coffee)
- Having milestones independent from conference deadlines
- Having regular stand-up meetings
- Coding together and reviewing each other's code (a lot)

All contributions should be recognised

Conceptualization

Mariona Coll Ardanuy^{1,5}
Daniel CS Wilson^{1,5}

Methodology

Mariona Coll Ardanuy
Federico Nanni¹
Kasra Hosseini¹

Implementation

Federico Nanni
Kasra Hosseini
Mariona Coll Ardanuy
Kaspar Beelen^{1,5}

Reproducibility

Kasra Hosseini
Federico Nanni

Interpretation

Kaspar Beelen
Mariona Coll Ardanuy
Katherine McDonough^{1,5}
Daniel CS Wilson
Ruth Ahnert⁵
Jon Lawrence⁴
Giorgia Tolfo²

Historical Analysis

Daniel CS Wilson
Katherine McDonough
Kaspar Beelen
Jon Lawrence

Data Curation

Kaspar Beelen
Mariona Coll Ardanuy
Federico Nanni
Giorgia Tolfo

Annotation

Giorgia Tolfo
Ruth Ahnert
Kaspar Beelen
Mariona Coll Ardanuy
Jon Lawrence
Katherine McDonough
Federico Nanni
Daniel CS Wilson

Writing and Editing

Mariona Coll Ardanuy
Federico Nanni
Ruth Ahnert
Kaspar Beelen
Kasra Hosseini
Jon Lawrence
Katherine McDonough
Barbara McGillivray^{1,3}
Daniel CS Wilson

Supervision

Barbara McGillivray
Ruth Ahnert

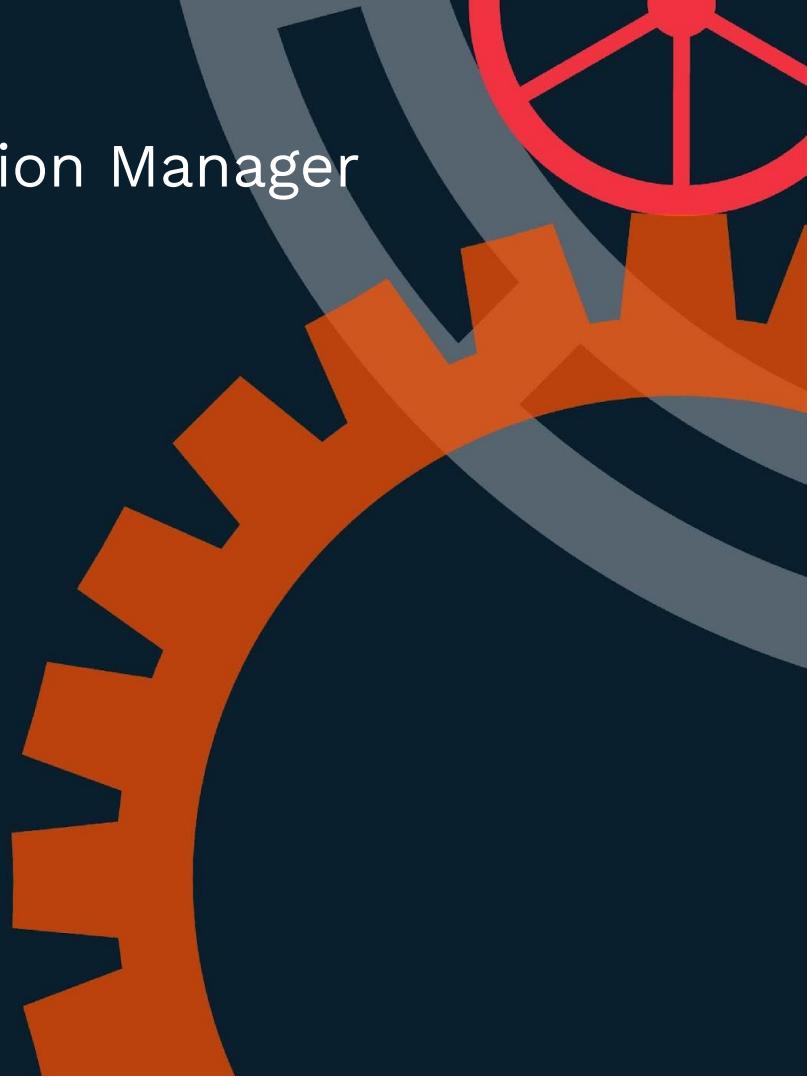
Project Management

Barbara McGillivray
Ruth Ahnert
Mariona Coll Ardanuy

Thank you! Questions?



Kalle Westerling, Research Application Manager
kwesterling@turing.ac.uk



Kalle Westerling, Research Application Manager kwesterling@turing.ac.uk



"Navigating Digital Humanities Careers Beyond the Ivory Tower" Formal Proposal for the ...

File Edit View Insert Format Tools Extensions Help

Formal Proposal for the Series Editors

Once the series editors have expressed interest in the idea for the volume, the next step is for the volume editors to prepare a formal proposal. This process allows the volume editors and series editors to ensure that their goals are aligned, and that the volume will be a good fit for the series. This process often involves several rounds of feedback and revision, resulting in a document that—with only minor modifications—becomes the proposal that is later submitted to the press for formal peer review.

Narrative Overview

A roughly 500-word overview of the volume, including how it will contribute to conversations currently underway in the digital humanities.

The digital humanities, as a discipline, sits at the intersection of traditional humanities scholarship and the digital world, embodying a unique blend of skills and perspectives. Despite its potential for wide-ranging applications, the conversation around DH careers has often been limited to academic positions. However, the dire reality of the job market and the evolution of industries outside academia demand a reevaluation of what constitutes a fulfilling and impactful career for DH scholars. Navigating Digital Humanities Careers Beyond the Ivory Tower emerges from this critical juncture in the humanities, where the decline in tenure-track positions and the rise of part-time and contingent roles necessitate explorations of alternative career paths that are both meaningful and viable for digital humanities (DH) scholars. Building upon the critical groundwork laid by Karina Rogers in *Putting the Humanities PhD to Work* (2010), this anthology aims to extend the discussion, advocating for a modified sense of professional success and fulfillment for humanists beyond the confines of academia.

This anthology contributes to this conversation by presenting a series of essays, case studies, and personal narratives that collectively argue for a more inclusive understanding of career possibilities. Our mission is twofold: to debunk the entrenched assumption that a tenure-track professorship is the sole pinnacle of career achievement for DH scholars, and to illustrate the wide array of opportunities that lie beyond. It offers a comprehensive exploration of the versatile career pathways available to scholars within and alongside the field of DH, as well as the narratives from those already successfully in professional roles outside of academia.

With contributions that not only share the challenges and triumphs of transitioning from academia to other sectors—ranging from for-profit industries and non-profit organizations to roles in libraries and government—but also those that highlight the impact of DH in transcending traditional humanities boundaries, this volume addresses

Kalle Westerling, Research Application Manager
kwesterling@turing.ac.uk

Who are you and what are you doing professionally?

What is your current position?

Kalle Westerling, PhD (He/they) • You
Open-source advocate for user-centric research software | Driving real-...
6d •

What parts of my career journey are you curious about? 🤔 I'm giving a talk next week on my path—from getting my PhD in the humanities to working in different research infrastructure roles like Research Software Engineer and Research Application Manager at The Alan Turing Institute (yep, the national hub for data science and AI! 🎉). What do you think I should talk about? Your suggestions would be super helpful! 🙏

#digitalhumanities #altac #careers #careerpaths #professionalisation
#researchinfrastructure #research

Shakir Laher and 44 others 7 comments • 2 reposts

Reactions

+37

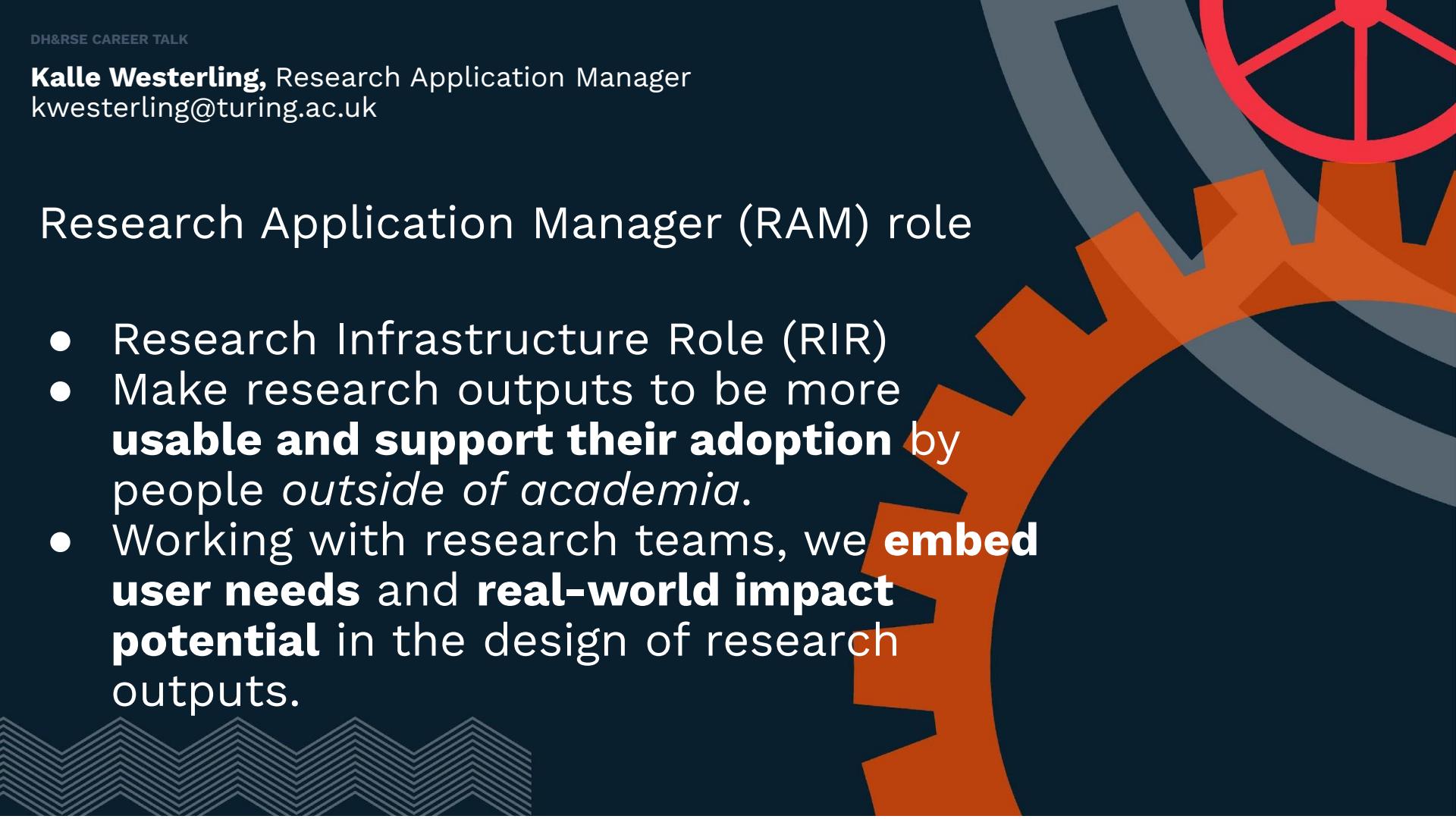
Like Comment Repost Send

2,366 impressions View analytics

Kalle Westerling, Research Application Manager
kwesterling@turing.ac.uk

Research Application Manager (RAM) role

- Research Infrastructure Role (RIR)
- Make research outputs to be more **usable and support their adoption** by people *outside of academia*.
- Working with research teams, we **embed user needs** and **real-world impact potential** in the design of research outputs.





Welcome

Welcome to *The Turing Way* handbook to reproducible, ethical and collaborative data science.

The Turing Way is an open science, open collaboration, and community-driven project. We involve and support a diverse community of contributors to make data science accessible, comprehensible and effective for everyone. Our goal is to provide all the information that researchers, data scientists, software engineers, policymakers, and other practitioners in academia, industry, government and the public sector need to ensure that the projects they work on are easy to reproduce and reuse.

Top Tip

The Turing Way is not meant to be read from start to finish. Start with a concept, tool or method that you need now, in your current work. Browse the different guides that make up the book, or use the search box to search for whatever you would like to learn about first.

All stakeholders from any level of experience or domain expertise are encouraged to use *The Turing Way* to understand their roles and responsibility of reproducibility in data science and research. You can inspect our resources on [GitHub](#), contribute to the project as described in our [contribution guidelines](#) and re-use all materials ([see the License](#)).

In the [Foreword Section](#) of this book, we provide meta-level information about *The Turing Way*, including guidance on how to use and navigate this project, how to cite the book and opportunities to get involved.



Legend

- Informational
- Formal
- Workshop
- Formal
- Informal
- Workshop

GOALS [Matrix of success]

STRETCH GOALS

Manager

Burndown charts

Core Task 1: Drive innovation in the development of data and software.

Core Task 2: Accelerate research innovation.

Core Task 3: Create sustainable communities.

Other Tasks: Advisory board, Lean inception.

TEA Product Compendium

This is a comprehensive document that aggregates all critical information regarding the TEA platform.

It is a living document, continually updated, and should be considered the central source of truth for everyone involved in the TEA product lifecycle.

It should not be considered a reference only for the development team and stakeholders but also as an engagement tool for potential investors and partners, showcasing the holistic view of the TEA platforms journey, its adherence to ethical standards, and its commitment to delivering value to users.

This document builds on previous work:

- The Trustworthy and Ethical Assurance (TEA) Platform
- TEA-DH
- Co-Working
- Phase 3 Planning

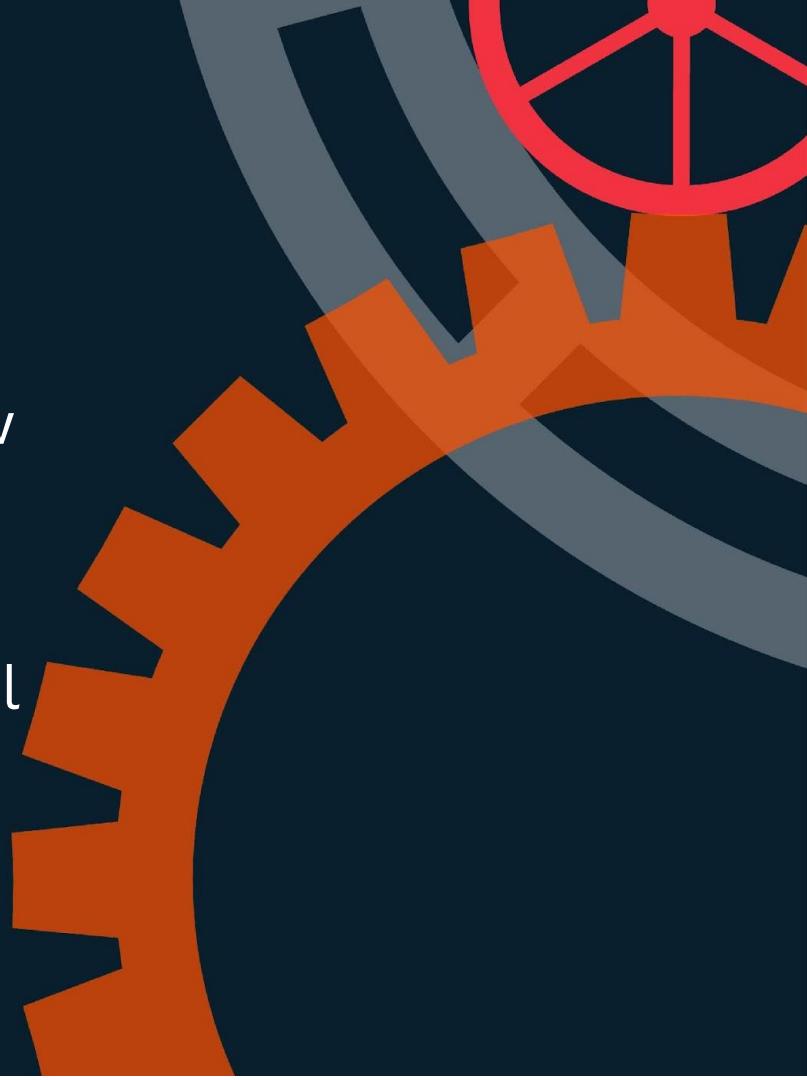
Executive Information about the Platform

- TEA Product Compendium
 - Executive Information about the Platform
 - Summary
 - Mission Statement
 - Vision Statement
 - Core Values
- Product Architecture
 - High-level architectural overview
- Product Roadmap
- Stakeholder map
- User Engagement
 - User Personas
 - Value Propositions
 - Quality of Compliance
 - Market Insights
 - Competitive Landscape Analysis
- Go-to-Market Strategy
 - Possible taglines

Kalle Westerling, Research Application Manager
kwesterling@turing.ac.uk

What were your past positions, how did you get to where you are now?

How were you trained in software development? Do you have a formal education?



Kalle Westerling, Research Application Manager

kwesterling@turing.ac.uk



Report accessible
via NEH's website

Lisa Marie Rhody and Kalle Westerling,
“Expanding Communities of Practice through DH
Andragogy”

The screenshot displays the NEH website's product page for the book "What We Teach When We Teach DH: Digital Humanities in the Classroom". The page includes the book's cover image, title, authors, and a brief description: "Exploring how DH shapes and is in turn shaped by the classroom". It also shows the price (\$35.00), formats (eBook, Hardcover, Paperback), and options to download the cover image or request an exam copy.

Popular repositories

- python: glob-DH-Curriculum Session on Python, a general purpose programming language used for a wide variety of computational tasks.
- curlie: glob-DH-Curriculum Session on curlie, a web crawler for tracking print files, learning on data, and collaborating with others. Includes curlie!
- command-line: glob-DH-Curriculum Session on command-line.
- databases: glob-DH-Curriculum Session on databases, including concepts such as relational databases and NoSQL.
- Python: glob-DH-Curriculum Session on Python, a general purpose programming language used for a wide variety of computational tasks.
- curlie: glob-DH-Curriculum Session on curlie, a web crawler for tracking print files, learning on data, and collaborating with others. Includes curlie!
- test-drive: glob-DH-Curriculum Session on test-drive, including discussion of cleaning data, creating test corpora, and analyzing test results.
- curlie: glob-DH-Curriculum Session on curlie, a web crawler for tracking print files, learning on data, and collaborating with others. Includes curlie!
- Repositories

Popular

People

Top languages

The screenshot shows several digital humanities websites:

- Nebraska Today**: A news portal for the University of Nebraska-Lincoln, featuring a section on "Research institute aims to bring diverse 'new storytellers' to digital scholars".
- DH@UVA**: The digital humanities portal for the University of Virginia, listing "About DH@UVA", "DH Certificate", "Pedagogy", "Workshops", and "Resources". It also features a "Workshop: The Triangle Digital Humanities Institute" page.
- dh+lib**: A platform for digital literacy, featuring a "Features" section and a "Digital Literacy on a Dime: Designing Community-Based Digital Humanities Institutes across North Carolina Universities" article.

GitHub
repositories



Dark mode

Fix stuck nodes

Community details

Explain visualization

• 521 / 522 • 3691 / 3691 • 0 • 57 • 11 / 12 • Q 60%

Settings

Co-occurrence Dataset
pre-filtered 14-day span

Node filtering
minDegree (0)

sizing
nodeMultiplier (60%)

node size from
Degree in current network

behavior
 auto-clear unconnected nodes
 stickyNodes

Detect communities
algorithm [Louvain]

Edge filtering
minWeight (0)

sizing
edgeMultiplier (100%)

weight from
Total number of co-appearing dates

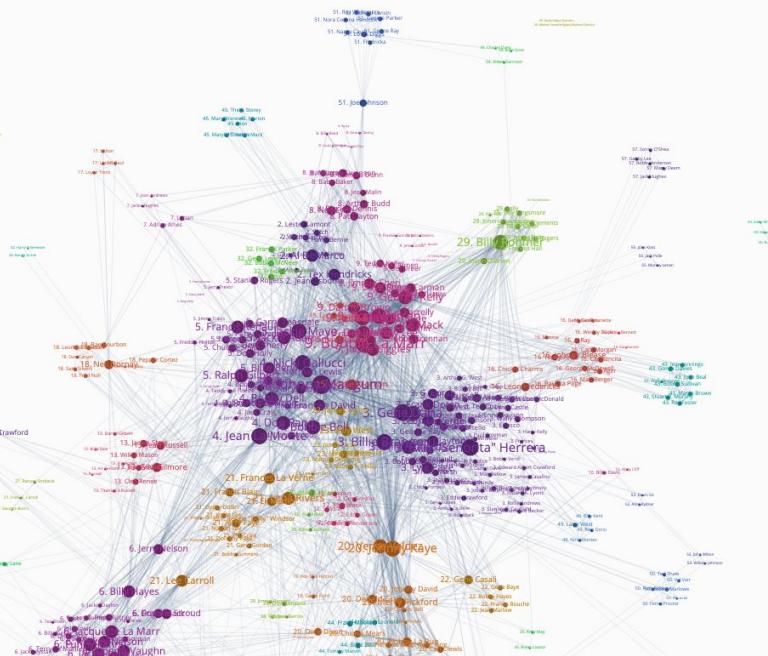
startYear 1930 endYear 1940

Force settings

Search

Dataset

Reset settings



PhD Dissertation project: Social network visualisation

Dataset: Drag performers in United States, 1930–1940

Stack: Vanilla JavaScript (?!)

Kalle Westerling, Research Application Manager

kwesterling@turing.ac.uk

Principal and Co-Investigators



Ruth Ahnert
(QMUL)



David Beavan
(Turing)



Emma Griffin
(UEA)

Project team



Claire Austin
Rights Assurance



Kaspar Beelen
Digital Humanities Senior
Research Associate



Mariona Coll Ardanuy
Computational Linguistics
Senior Research Associate



André Piza
Research Project Manager



Griffith Rees
Research Data Scientist



Kalle Westerling
Research Software Engineer



Timothy Hobson
(Turing)



Jon Lawrence
(Exeter)



Maja Maricevic
(British Library)



Lélén Demertzí
Programme Coordinator



Luke Hare
Research Data Scientist



Sherman Lo
Research Data Scientist



Daniel Wilson
History Senior Research
Associate



Rosie Wood
Research Data Scientist



Barbara McGillivray
(Turing / King's College
London)



Mia Ridge
(British Library)



Alan Wilson
(Turing)



Katie McDonough
Senior Research Associate



Federico Nanni
Senior Research Data
Scientist



Nilo Pedrazzini
Corpus-Based Digital
Humanities Research
Assistant

Kalle Westerling, Research Application Manager

kwesterling@turing.ac.uk

Date: Sat, 30 Oct 2021 16:41:38 +0000
From: Code4Lib Jobs <jobs.code4lib.org/>
Subject: Job: Digital Humanities research software engineer at British Library / Living with Machines / Alan Turing Institute

We really want to reach a diverse range of people with this job ad. Don't be put off by the title - other relevant titles include: software developer, creative technologist, front-end developer, data scientist, data or visualisation specialist, or digital humanities researcher who uses computational methods.

Find out more and apply: https://britishlibraryrecruitmentcells.com/brit/gapea/vacancy_id/1002678

The post is full time until July 2023, the salary range is £39 - 47k, and it can be remote 4 days a week, with one day a week at the British Library in London (Boston Spa possibility too).

Job titles are tricky, and job descriptions make for dry reading, so I interviewed the wonderful outgoing post holder Olivia Vane about her experience in the role: <https://livingwithmachines.ac.uk/what-does-a-digital-humanities-research-software-engineer-do/>

Applications close on November 7th - that's next Sunday - so please pass this on or take a look soon!

If you have any questions, please email digitalresearch@bl.uk (or use the email address on <https://bl.uk/digital> if that's blocked).

Full Time, Fixed Term till 31 July 2023

Living with Machines (LwM) is an ambitious large-scale project in data science and the digital humanities. Researchers from a range of disciplines - including historians, engineers, data scientists, geographers, computational linguists, library professionals and curators – are working together to create research methods, tools and data. We challenge assumptions about how different disciplines can interact and break down academic traditions.

We are looking for a Digital Humanities Research Software Engineer (DH RSE) to complement our team and create spatial and temporal representations of complex histories and interfaces for specific research methods and technologies. This is an exciting opportunity to contribute to the development and implementation of the digital scholarship outreach streams of the LwM project by assembling, designing, implementing, developing and integrating a range of tools.

Working as part of a multi-disciplinary team, the successful candidate will help identify requirements, design and implement online interfaces that integrate different parts of the project. They will support the collections, questions and methods of the project. The DH RSE will create outputs including creative, intuitive visualisations and interfaces for digitised collections, crowdsourcing tasks and data science outputs for project specialists, academic and public users.

The post holder will be responsible for developing and maintaining the interface for the community of practice. This includes training, academic or other appropriate environments. To work as a Research Software Engineer, software developer, creative technologist, data or visualisation specialist or digital humanities researcher. They will have excellent technical skills, including experience of the tools and technologies that support digital scholarship. Excellent oral and written communication skills are also essential for other Research Software Engineers (<https://rse.ac.uk/>), posts; the post holder will have the opportunity to develop their skills and play an active part in all aspects of research including analysis and publication.

More about Living with Machines

This project aims to use computational techniques and large-scale datasets in order to ask questions about the ways in which technology altered the very fabric of human existence on a hitherto unprecedented scale. The project exploits a corpus of digitised sources, including newspapers, trade directories, census data, and other resources. By developing intuitive computational interfaces and a philosophy of interdisciplinary collaboration we will enable close interaction between computational methods and historical inquiry.

More about the British Library

We welcome and encourage job applications from people of all backgrounds. We particularly welcome applications from disabled and Black, Asian and Minority Ethnic (BAME) candidates as BAME and disabled people are currently under-represented throughout the British Library at this level.

In return we offer a competitive salary and a number of excellent benefits. Our pension scheme is one of the most valuable benefits we offer, as our staff can become members of the Alpha Pension Scheme where the Library contributes 20.9%. Another significant benefit the Library provides is the provision of a flexible working hours scheme which could allow you to work your hours flexibly over the week and to take up to 5 days flexi leave in a 3 month period. This is on top of 25 days holiday from entry, and public and privilege holidays.

As one of the world's great libraries, our duty is to preserve the nation's intellectual memory for the future. At the moment we have well over 150 million items, in most known languages, with three million new items added every year. We have manuscripts, maps, newspapers, magazines, prints and drawings, music scores, and patents. We operate the world's largest document delivery service providing millions of items a year to customers all over the world. What matters to us is that we preserve the national memory and enable knowledge to be created both now and in the future.

For further information and to apply, please visit www.bl.uk/careers quoting vacancy ref: 03841

Closing Date: 7th November 2021

Interview Date: 17, 19 November 2021

Brought to you by code4lib jobs: <https://jobs.code4lib.org/jobs/50202-digital-humanities-research-software-engineer>

“We really want to reach a diverse range of people with this job ad. Don't be put off by the title - other relevant titles include: software developer, creative technologist, front-end developer and JavaScript developer, data scientist, data or visualisation specialist, or digital humanities researcher who uses computational methods.”



Katie McDonough
Senior Research Associate



Federico Nanni
Senior Research Data
Scientist



Nilo Pedrazzini
Corpus-Based Digital
Humanities Research
Assistant



Kalle Westerling
Research Software Engineer

Kalle Westerling, Research Application Manager

kwesterling@turing.ac.uk

Principal and Co-Investigators

Ruth Ahnert (QMUL)	David Beavan (Turing)	Emily (UEA)

Timothy Hobson (Turing)	Jon I (Exeter)	Lorraine (Lancaster)

Barbara McGillivray (Turing / King's College London)	Mia I (British Library)	Michael (Oxford)



The slide displays several screenshots of GitHub repositories related to the project:

- zoonyper** (Public): A repository for processing Zooniverse annotations in Python/Jupyter Notebooks. It includes documentation, workflows, and various scripts like `update_all_contributors.py` and `reorganise_documentation.py`.
- label-studio-converter** (Public): A repository for creating ready-to-use Label Studio pre-populated JSON files from popular OCR formats. It contains commits from Kalle Westerling and others, including `pyseresent` and `transkribus` converters.
- LwM-BL-OS-maps** (Private): A repository for digitising survey maps from the British Library. It includes a README file and a Jupyter notebook titled "Machines Digitisation of Survey Maps.ipynb".
- alto-model** (Private): A repository for renaming XPaths in XML files. It contains commits from Kalle Westerling and others, including `dist`, `src/aldo_model`, and `tests` branches.

Steps to get Defoe + Jupyter Notebook running in current development status...

Requirements:

- An Azure VM
 - Mine is set up as a Standard D8s v3 (8 vcpus, 32 GB memory). Not needed to keep something this size running – but bigger = faster most of the time here...
 - You'll also probably want to set up a data disk and mount it to your VM.
 - Installed software: blobfuse (if you want to mount blobs, see next point), Spark + pypark setup incl. Java. (see below for install of software)
 - Since you'll most likely want to use a notebook connection to Defoe/querying, you'll also want to make sure your correct IP ports are opened for traffic, but for safety, I've added a password and restricted access to only certain IP addresses. Note: When you're on the Turing wifi, you won't be able to access the VM unless it whitelists the IP address.
- A mount point where your newspaper XML files are mounted (or a local version of the newspaper collections needed)
 - SAS token for the Azure container
 - Shell script for mounting (I use a script which is attached below, see "Bash script for mounting lwm and HMD collections")
- A list of all the relevant filepaths for those XML files (created using `get_filepaths` function described in detail below)
 - (You might also want a general understanding of the structure of the "API" of FMP documents) — see "API Structure" below

The following screenshots will guide you through the set up of an Azure virtual machine and the current development branch of Defoe for the Living with Machines project.

Kalle Westerling, Research Application Manager

kwesterling@turing.ac.uk

The screenshot shows the official website of The Alan Turing Institute. At the top, there's a navigation bar with links for Home, Events, News, About us, Research, Skills, People, and Opportunities. Below this, a teal-colored header bar contains links for TRIC-DT, Research themes, People, Community, Projects and partners, News and media, and Contact us. The main content area features a large image of a digital twin simulation. Overlaid on the image is the title "Turing Research and Innovation Cluster in Digital Twins". Below the title are two buttons: "Learn more" and "Follow us on twitter".

Introduction

A pioneering new initiative in Digital Twins research and innovation.

The TRIC-DT mission is to carry out fundamental research relating to digital twins that addresses the most urgent challenges facing society and achieves real-world impact.

Digital twins offer exciting new possibilities across a wide range of sectors from health, environment, transport, manufacturing, defence, and infrastructure. By connecting the virtual and physical worlds, digital twins can better support decisions, extend the operational life of systems and services, explore different scenarios, and introduce many other efficiencies and benefits. As a result, digital twins have been identified by governments, professional bodies, and industry, as a key technology to help address many of the societal challenges we face.

Digital twins will play a crucial role in delivering solutions to the global-scale challenges, such as climate change, which currently face humanity. While DT technology has proven to be extremely powerful in a range of areas, current DTs are often highly bespoke, and their design, development, and deployment to address real-world problems often requires specialised expertise and computational infrastructure, which can create a barrier that limits their scalable use. The TRIC-DT will work with key partners in an open and reproducible manner to democratise access to open, reproducible and trustworthy ecosystems of digital twins.



This screenshot shows a GitHub repository page for "alan-turing-institute / AssurancePlatform". The repository has 26 branches and 95 issues. The codebase includes files like README, CONTRIBUTING.md, and various configuration and Docker files. A "Get started" button is visible at the bottom left of the repository view.

The screenshot shows the homepage of the Trustworthy and Ethical Assurance (TEA) Platform. It features a large illustration of people working together around a table with digital screens. The main heading is "Build trust, collaboratively." Below the heading, a brief description states: "The Trustworthy and Ethical Assurance (TEA) Platform is an innovative, open-source tool designed to facilitate the process of creating, managing and sharing assurance cases for data-driven technologies, such as digital twins or AI." There are also "Documentation" and "GitHub" links at the top.

This screenshot shows a detailed view of the TEA Platform's user interface. It displays an "assurance case" flowchart. The flow starts with a pink rounded rectangle labeled "G1 A goal set in our assurance process". This leads to a blue rounded rectangle labeled "C1 Here's the project's context". From "C1", the flow continues to a blue rounded rectangle labeled "P1 This is a property claim". From "P1", it branches to another blue rounded rectangle labeled "P2 This is also a property claim". Finally, the flow ends at a green rounded rectangle labeled "E1 Here's evidence". The interface includes a search bar, a "Resources" dropdown, and a "Feedback" link at the top right.

Kalle Westerling, Research Application Manager

kwesterling@turing.ac.uk

The screenshot shows the Data/Culture website. At the top, there's a navigation bar with links for Home, Events, News, About us, Research, Skills, People, and a search bar. Below this is a banner with the text "Data/Culture: Building sustainable communities around Arts and Humanities datasets and tools". A "Learn more" button is visible. The main content area has a dark background with white text. It starts with an "Introduction" section, followed by a paragraph about the project's purpose and a list of its pilot year goals. There's also a sidebar with links to Project aims, Applications, How can you get involved?, Related content, Get in touch, Organisers, and Researchers and collaborators.

This screenshot shows two sections from the Data/Culture website: "Organisers" and "Researchers and collaborators". Both sections feature a grid of profiles for various team members, each with a small photo, name, title, and a brief description. The "Organisers" section includes Dr Pieter Francois, Professor Ruth Ahnert, Dr Katherine McDonough, David Beavan, Dr Timothy Hobson, and Dr Federico Nanni. The "Researchers and collaborators" section includes Dr Pieter Francois, Dr Katherine McDonough, Professor Ruth Ahnert, David Beavan, Dr Federico Nanni, Ed Chalstrey, Dr Daniel Wilson, Dr Nilo Pedrazzini, and Dr Kalle Westerling.

Two parts:
 - Living with Machines
 - Seshat

Introduction

Data/Culture is a sandbox for the re-use of data and tools in the humanities and arts, in ways that develop high-quality research collaborative communities.

In its pilot year (2023-24), the team is focused on:

- Building communities of historians around 1) the *Seshat: Global History Databank* and 2) the *Living with Machines* project, whose outcomes include the MapReader software library, as well as data and tools related to historical British newspapers.
- Establishing a stronger network of Research Software Engineers for the Arts and Humanities, driving collaborative innovation and embedding skills. Delivering a roadmap for a national capability tailored to Arts and Humanities needs.

Explaining the science

Data/Culture is a unique experiment in arts and humanities research—it provides extended support for projects that have invested expertise, time, and of course funding in the development of tools and data for digital research. At the same time, it is an environment in which to scope the needs of digital research infrastructure in the arts and humanities so that future investment in these areas has a longer lifespan and leads to significant research findings across these disciplines.

Due to the nature of the current project-focused funding landscape, tools and data often never find an audience beyond the original project team, let alone a wider community. Even with the best intentions, new projects often reinvent the wheel, building the same or very similar datasets and tools, because they do not know about, or cannot access, open research components that already exist. Other times, projects develop data that is so bespoke, it is nearly impossible to re-use.

If those components could be designed with re-usability in mind up front, if they were well packaged and documented, if communities of users and maintainers were actively built around them, and if skills to exploit them were embedded in disciplinary pedagogy, we could create the basic components of an open and modular digital research infrastructure that would help accelerate research innovation.

This infrastructure could take many forms – however, knowledge about what works specifically for the arts and humanities is sparse and anecdotal. What is needed is more extensive work to define standards and models of good practice. Examples from initiatives outside the UK (such as DARIAH across Europe, Huma-Num in France, or CLARIAH in the Netherlands) as well as efforts within the UK (e.g. Towards a National Collection), will collaborate with colleagues to shape a future where the arts and humanities are as well-placed to pursue digital research as other disciplines.

Data/Culture is therefore a space in which to demonstrate what happens when we build and re-use data and tools with the aim of contributing to digital infrastructure for the arts and humanities.

Kalle Westerling, Research Application Manager

kwesterling@turing.ac.uk

Seshat-Global-History-Databank / seshat_api

Type / to search

seshat Public forked from [edwardshastrey/seshat](#)

Type / to search

seshat Public forked from [edwardshastrey/seshat](#)

Type / to search

Seshat API

Caution

This package is currently actively working on building or subject to significant changes.

Please do not use this package if it contains essential features, and users should proceed with caution.

Thank you for your understanding.

This is the Python binding for Seshat database. The Seshat library is used to query the database and interact with it.

Installation

Seshat: Global History Databank

Seshat was founded in 2011 to bring together the most current and comprehensive body of knowledge about human history in one place.

This repo contains the necessary Django Python code to host the [Seshat](#) website and interact with the backend PostgreSQL database.

Developers

Follow the instructions available in [docs/source/getting-started/setup/index.rst](#).

Living-with-machines / alto2txt

Type / to search

MapReader Public

A computer vision pipeline for exploring and analyzing images at scale

MapReader Public

A computer vision pipeline for exploring and analyzing images at scale

Living-with-machines / T-Res

Type / to search

T-Res Public

Your dev branch isn't protected

Protect this branch

Living-with-machines / T-Res

Type / to search

T-Res: A Toponym Resolution Pipeline for Digitised Historical Newspapers

Overview

T-Res is an end-to-end pipeline for toponym detection, linking, and resolution on digitised historical newspapers. Given an input text, T-Res identifies the places that are mentioned in it, links them to their corresponding Wikidata IDs, and provides their geographic coordinates. T-Res has been developed to assist researchers explore large collections of digitised historical newspapers, and has been designed to tackle common problems often found when dealing with this type of data.

The pipeline has three main components:

- The Recogniser performs entity recognition.
- The Ranker performs candidate selection and ranking.
- The Linker performs entity linking and resolution.

The three components are used in combination in the Pipeline class.

DeezyMatch Public

A Flexible Deep Neural Approach to Fuzzy String Matching

DeezyMatch Public

A Flexible Deep Neural Approach to Fuzzy String Matching

DeezyMatch Public

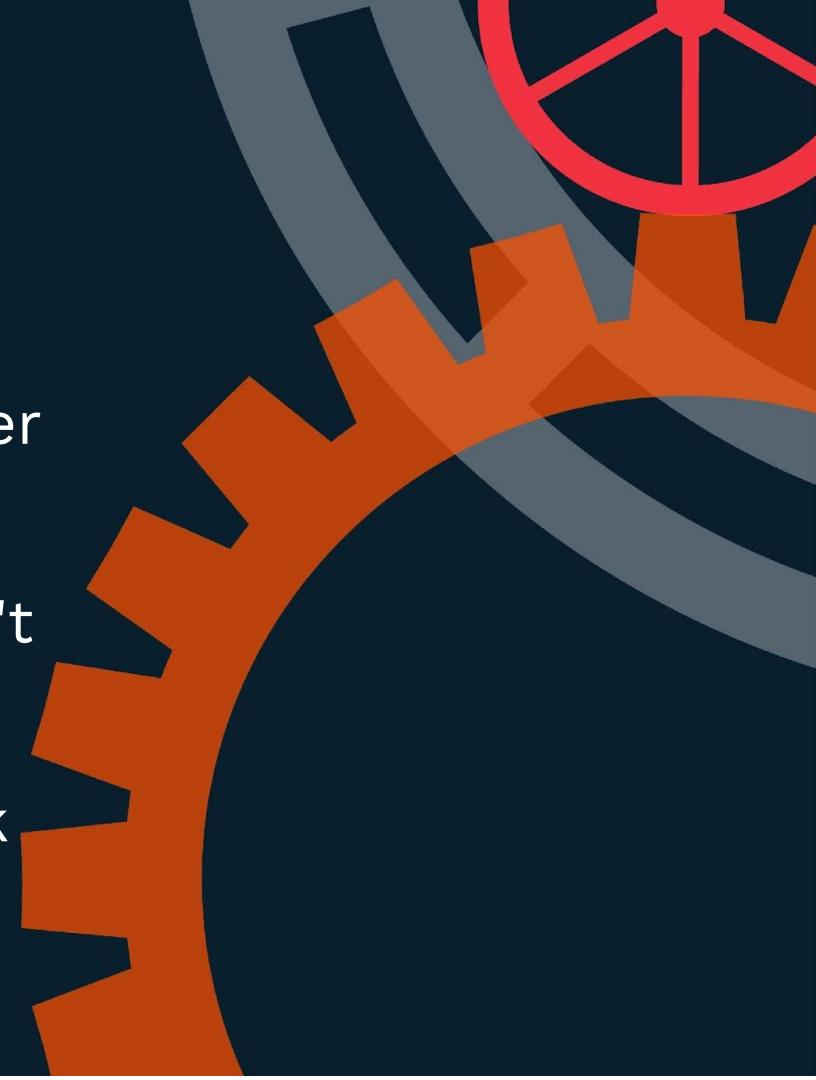
A Flexible Deep Neural Approach to Fuzzy String Matching

Kalle Westerling, Research Application Manager
kwesterling@turing.ac.uk

How were the first few months after
“making the jump”?

Was there any technology you didn't
get right away?

What resources did you use to pick
up your skills? Did teaching play a
role?



Kalle Westerling, Research Application Manager
kwesterling@turing.ac.uk

pluralsight.com

The screenshot shows the Pluralsight Skills dashboard. On the left, a sidebar menu includes Home, Browse all, Iris, Channels, Role IQ, Certifications, Paths, Skill IQ, Help & support, and Send Feedback. The main area displays a "Continue learning" section with three courses: "Next.js 13 Fundamentals" (Resume, 51m remaining), "Getting Started with Pulumi" (46m remaining), and "Terraform - Getting Started" (3h 35m remaining). Below this is a "Recommended courses to watch next" section featuring "React 18: The Big Picture" (1h 6m), "ASP.NET Core Web API Fundamentals" (8h 34m), and "Python 3 Fundamentals" (2h 55m). To the right, a profile summary for Kalle Westerling (Node Web Developer) shows a weekly goal of 21 min / 2 hours completed this week. The "Your badges" section highlights the "Month to Month 2x" badge (Get this badge and new skills) and lists upcoming badges: "Next.js 13 Fundamentals", "Skill IQ 5x", and "Scholarly Pursuit".

Kalle Westerling, Research Application Manager
kwesterling@turing.ac.uk

What kind of advancement opportunities/career path options are there at your current institution?

How to keep one foot in the canon/human condition and another on the rocket ship?

Formal Proposal for the Series Editors

Once the series editors have expressed interest in the idea for the volume, they will work with the volume editor to prepare a formal proposal. This process allows the volume editor to refine the scope and series editors to ensure that their goals are aligned, and that the volume fits the needs of the series. This process often involves several rounds of feedback and refinement, resulting in a document that—with only minor modifications—becomes the proposal for formal peer review.

Narrative Overview

A roughly 500-word overview of the volume, including how it will contribute to conversations currently underway in the digital humanities.

The digital humanities, as a discipline, sits at the intersection of traditional scholarship and the digital world, embodying a unique blend of skills and perspectives. Despite its potential for wide-ranging applications, the conversation around DH has often been limited to academic positions. However, the diverse reality of DH and the evolution of industries outside academia demand a reevaluation of what constitutes a fulfilling and impactful career for DH scholars. Navigating DH Humanities Careers Beyond the Ivory Tower emerges from this critical conversation. In the field of DH, where the decline in tenure-track positions and the rise of part-time and contingent roles necessitate explorations of alternative career paths that are meaningful and viable for digital humanities (DH) scholars. Building upon the groundbreaking work laid by Karina Rogers in Putting the Humanities PhD to Work, this anthology aims to extend the discussion, advocating for a redefined vision of success and fulfillment for humanities beyond the confines of academia.

This anthology contributes to this conversation by presenting a series of essays, interviews, and personal narratives that collectively argue for a more inclusive understanding of career possibilities. Our mission is twofold: to debunk the assumption that a tenure-track professorship is the sole pinnacle of career success for DH scholars, and to illuminate the wide array of opportunities that lie ahead. The book offers a comprehensive exploration of the versatile career pathways available within and alongside the field of DH, as well as the narratives from those successfully in professional roles outside of academia.

With contributions that not only share the challenges and triumphs of transitioning from academia to other sectors—ranging from for-profit industries and non-profits to government—but also those that highlight the significant impact of DH in reexamining traditional humanities boundaries, this volume

Kalle Westerling, Research Application Manager
kwesterling@turing.ac.uk

What kind of advancement opportunities/career path options are there at your current institution?

How to keep one foot in the canon/human condition and another on the rocket ship?

"Navigating Digital Humanities Careers Beyond the Ivory Tower" Formal Proposal for the ...

File Edit View Insert Format Tools Extensions Help

100% | Heading 4 | Droid ... | - 11 + | B I U A

Formal Proposal for the Series Editors

Once the series editors have expressed interest in the idea for the volume, the next step is for the volume editors to prepare a formal proposal. This process allows the volume editors and series editors to ensure that their goals are aligned, and that the volume will be a good fit for the series. This process often involves several rounds of feedback and revision, resulting in a document that—with only minor modifications—becomes the proposal that is later submitted to the press for formal peer review.

Narrative Overview

A roughly 500-word overview of the volume, including how it will contribute to conversations currently underway in the digital humanities.

The digital humanities, as a discipline, sits at the intersection of traditional humanities scholarship and the digital world, embodying a unique blend of skills and perspectives. Despite its potential for wide-ranging applications, the conversation around DH careers has often been limited to academic positions. However, the dire reality of the job market and the evolution of industries outside academia demand a reevaluation of what constitutes a fulfilling and impactful career for DH scholars. Navigating Digital Humanities Careers Beyond the Ivory Tower emerges from this critical juncture in the humanities, where the decline in tenure-track positions and the rise of part-time and contingent roles necessitate explorations of alternative career paths that are both meaningful and viable for digital humanities (DH) scholars. Building upon the critical groundwork laid by Katrina Rogers in *Putting the Humanities PhD to Work* (2010), this anthology aims to extend the discussion, advocating for a redefined sense of professional success and fulfillment for humans beyond the confines of academia.

This anthology contributes to this conversation by presenting a series of essays, case studies, and personal narratives that collectively argue for a more inclusive understanding of career possibilities. Our mission is twofold: to debunk the entrenched assumption that a tenure-track professorship is the sole pinnacle of career achievement for DH scholars, and to illuminate the wide array of opportunities that lie beyond. It offers a comprehensive exploration of the versatile career pathways available to scholars within and alongside the field of DH, as well as the narratives from those already successfully in professional roles outside of academia.

With contributions that not only share the challenges and triumphs of transitioning from academia to other sectors—ranging from for-profit industries and non-profit organizations to roles in libraries and government—but also those that highlight the impact of DH in transcending traditional humanistic boundaries, this volume addresses