

INTRODUCTION TO DIGITAL HUMANITIES

PRINCIPLES OF DIGITAL HUMANITIES II

Dr. Thomas Köntges,
Digital Humanities, University of Leipzig

10 November 2015

REPETITION

HOT APPROACHES IN THE HUMANITIES

Colonialist Bias

Gender studies

Deconstructing Knowledge

New Philology

= De-biasing

RESEARCH QUESTION

Guides your research

Clear and focused

Collect Data that will answer the research question

Watch bias in your data/question

OTHER POTENTIAL BIAS IN RESEARCH

Data Bias (sampling or scope)

Visualisation Bias

Statistical/Algorithm-based Bias (related to visualisation bias)

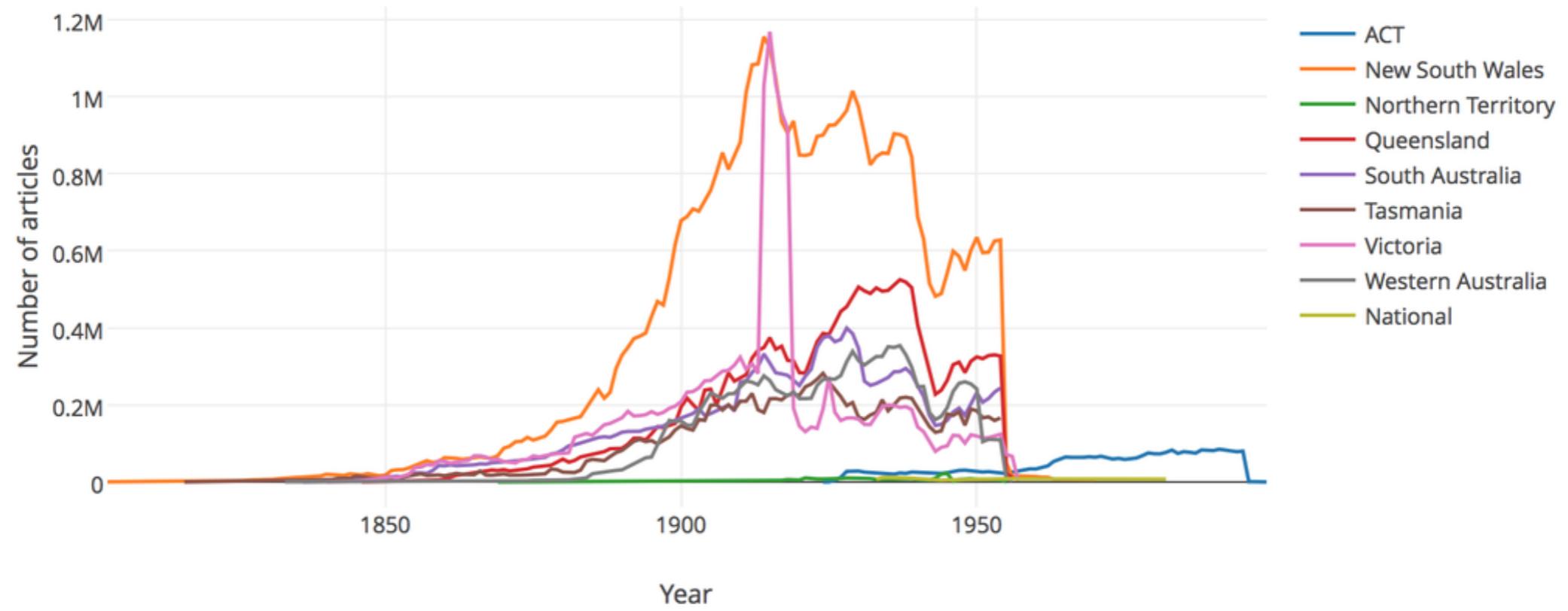
Reducing **all** Bias in Research

Disclose Bias if it cannot be avoided

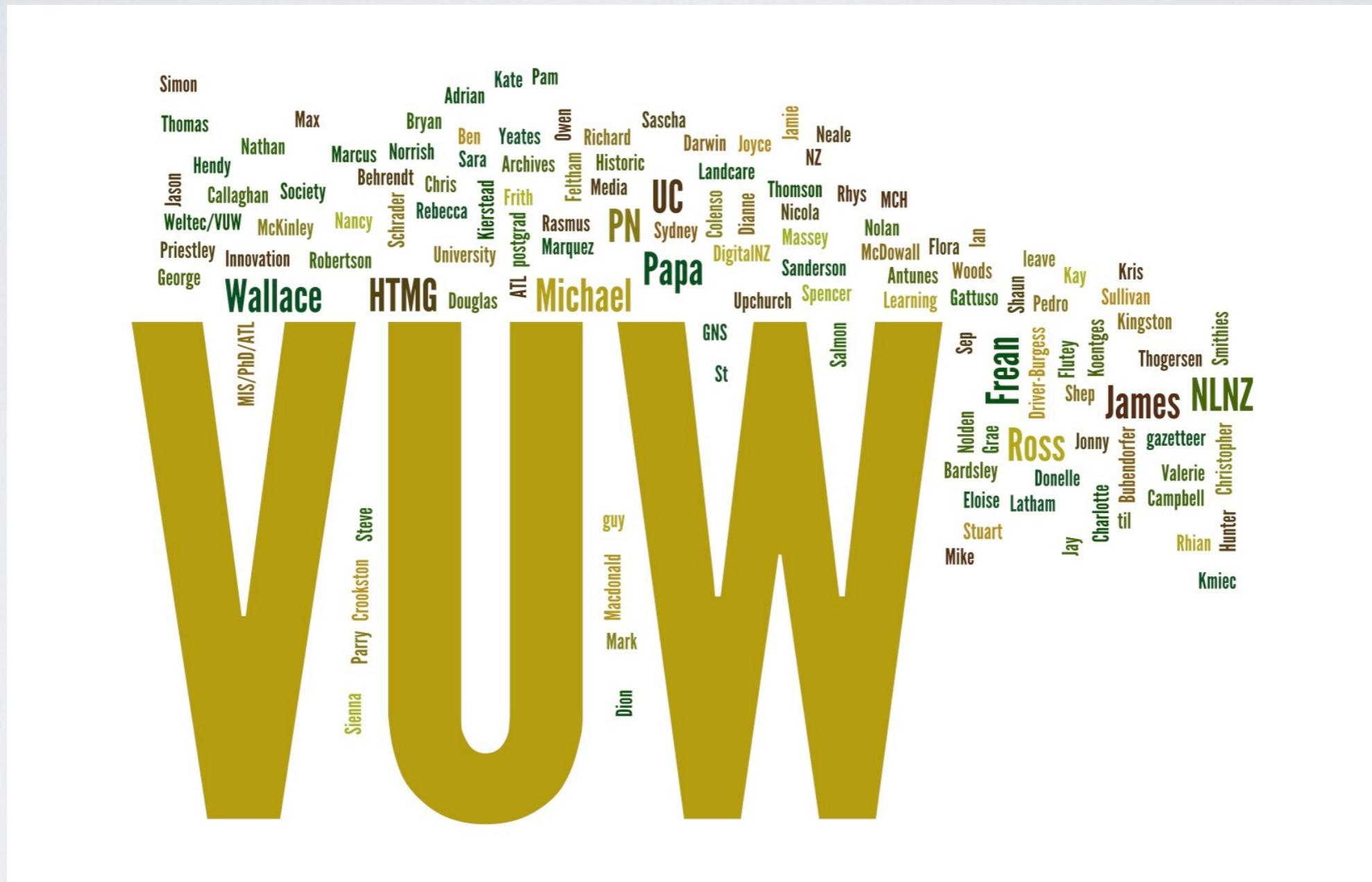
DATA BIAS



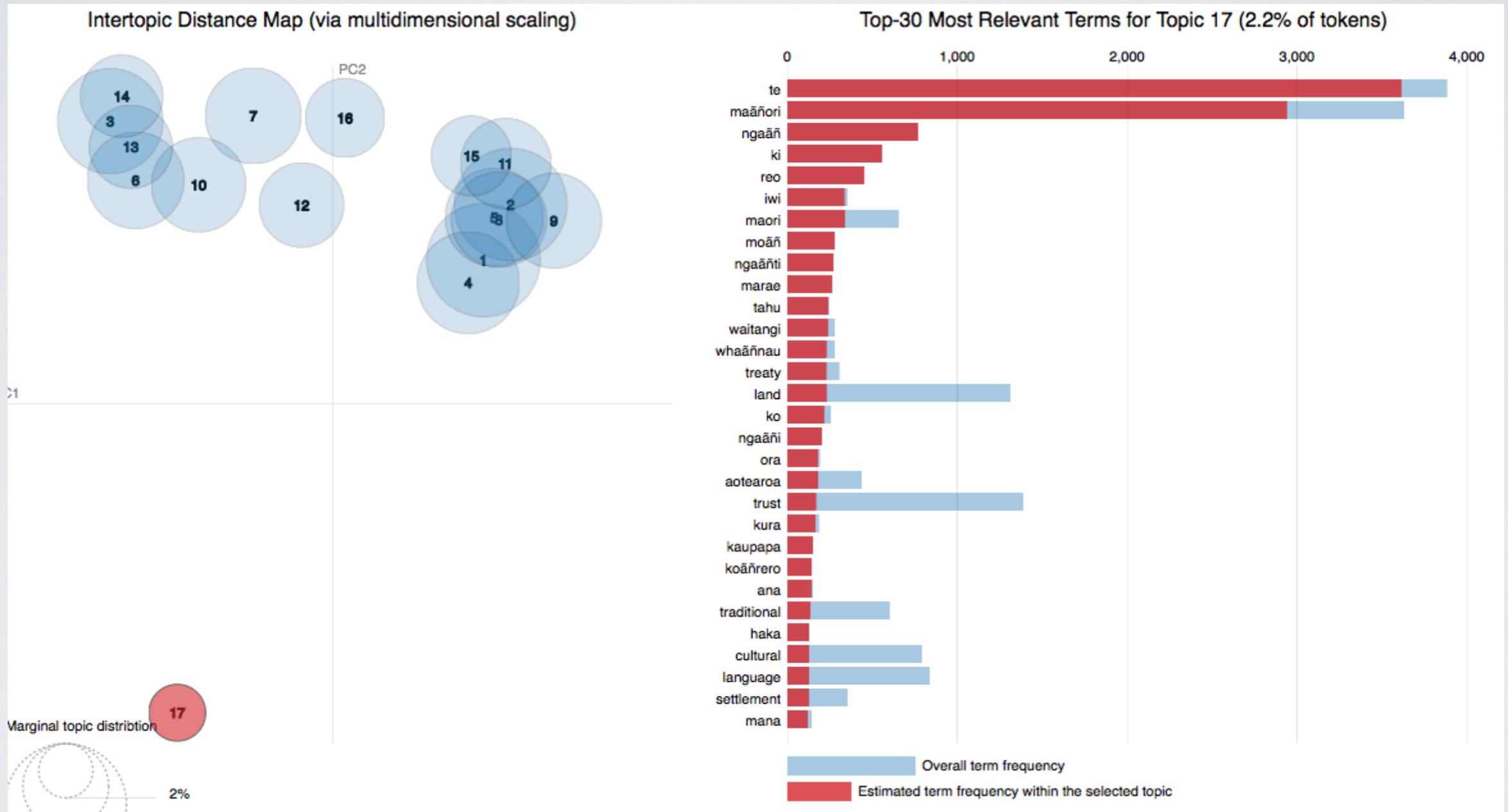
Trove newspaper articles by state



VISUALISATION BIAS



ALGORITHM-BASED BIAS



BE A GOOD DH-RESEARCHER

Research Question is central

Reduce **all** Bias in Research

Disclose Bias if it cannot be avoided

Explain Method (There should not be a magic box)

Produce Reproducible Results

Publish Openly and publish your Data too (if possible)

RESEARCH ON DIGITAL AVATARS

Thomas Köntges
Introduction to Digital Humanities
University of Leipzig



Digital Humanities
UNIVERSITÄT LEIPZIG

DIGITISATION

- Increase Accessibility of Research Data
- OCR
- Production of Digital Avatars
- **Not** lossless

EXAMPLE: SPECIALISED METHODS

COMPUTER AIDED TEXT ANALYSIS

Thomas Köntges
Introduction to Digital Humanities
University of Leipzig



Digital Humanities
UNIVERSITÄT LEIPZIG

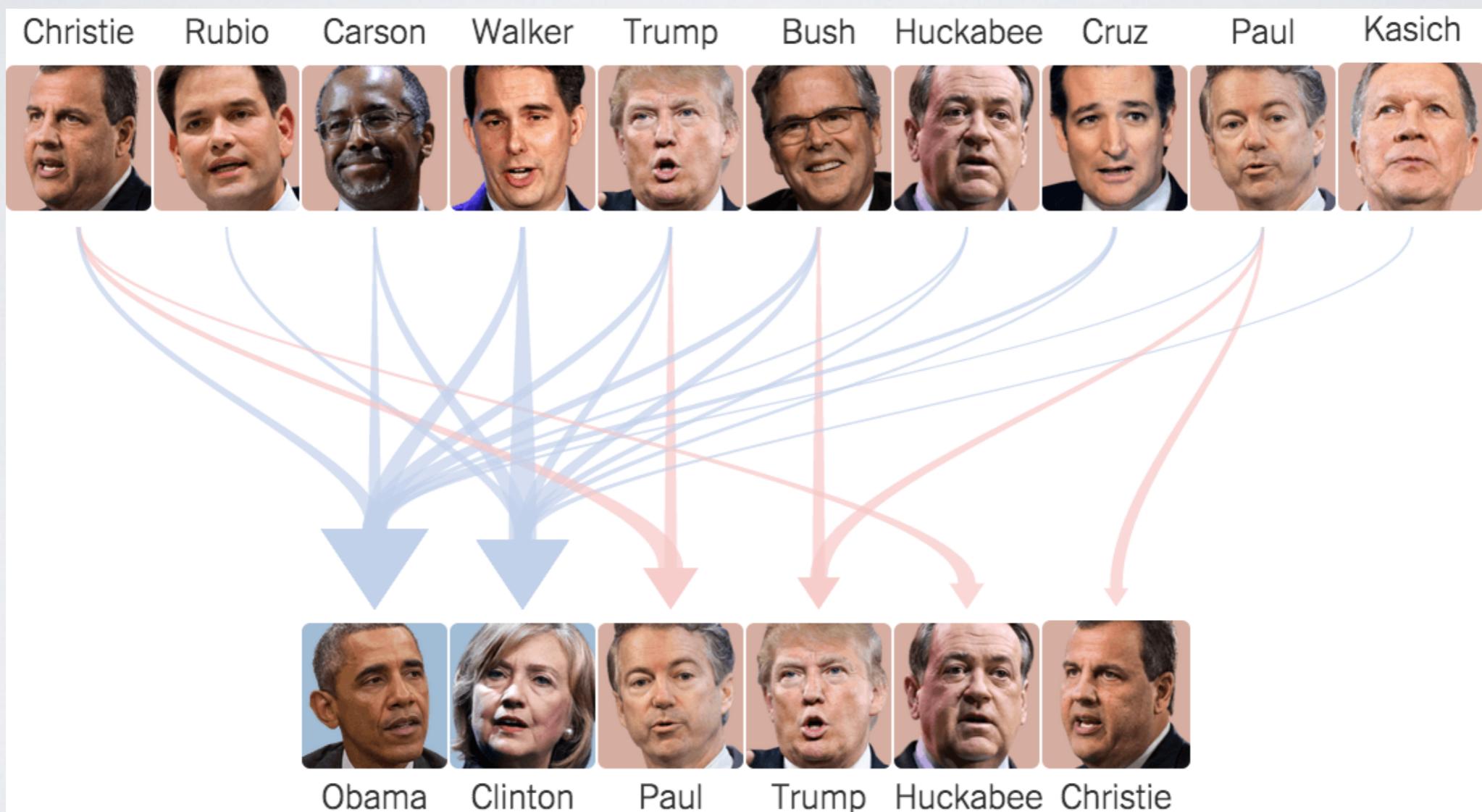
COMPUTER AIDED TEXT ANALYSIS

- **Analysis:** Split the texts into basic units (e.g. morphemes, words)
- **Synthesis:** Reassemble those units to a new text

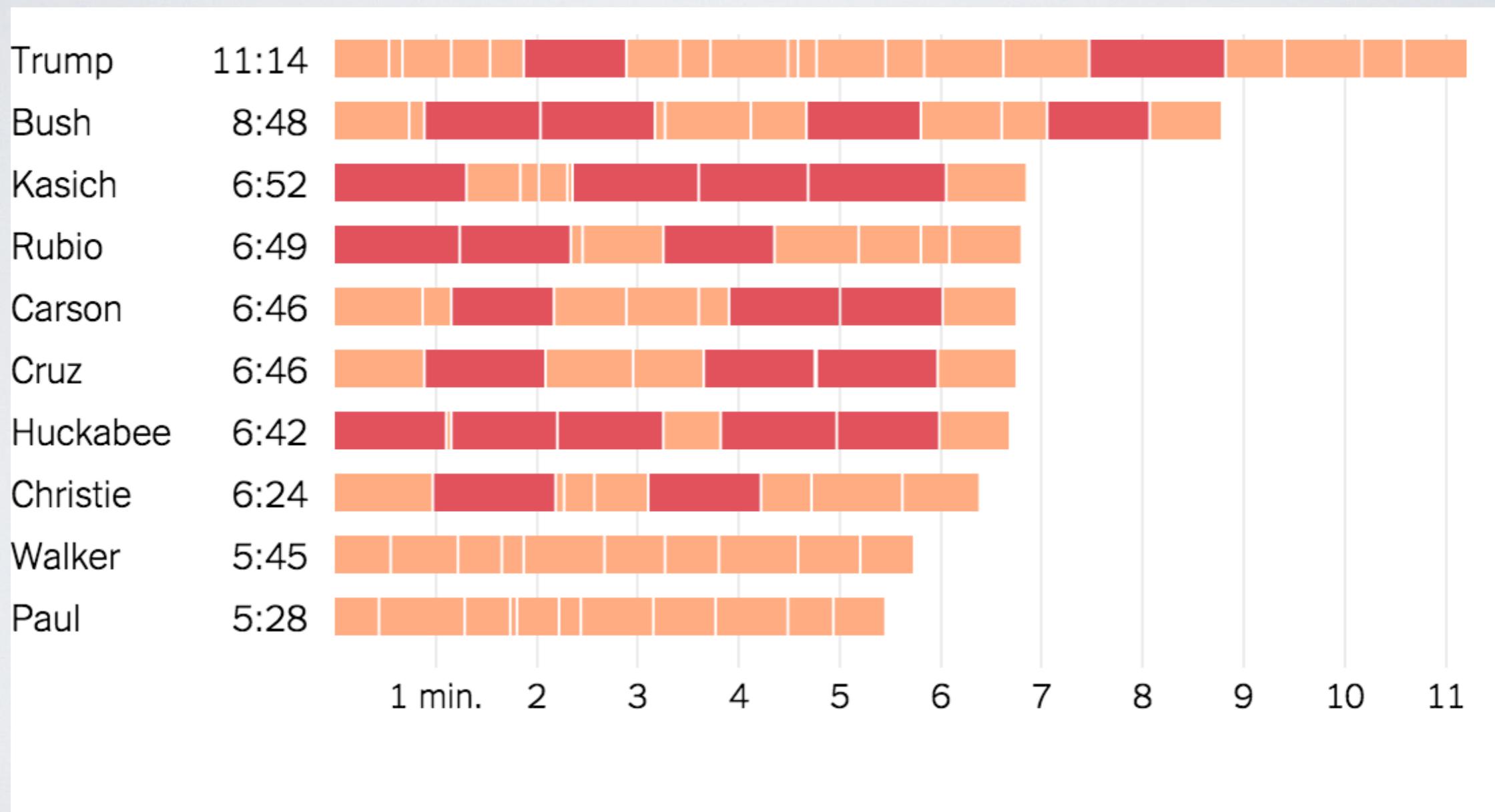
TEXT ANALYSIS SYSTEMS

- Search large texts/corpora quickly
- Complex searches
- Visualise the results, so we better understand the text/corpora

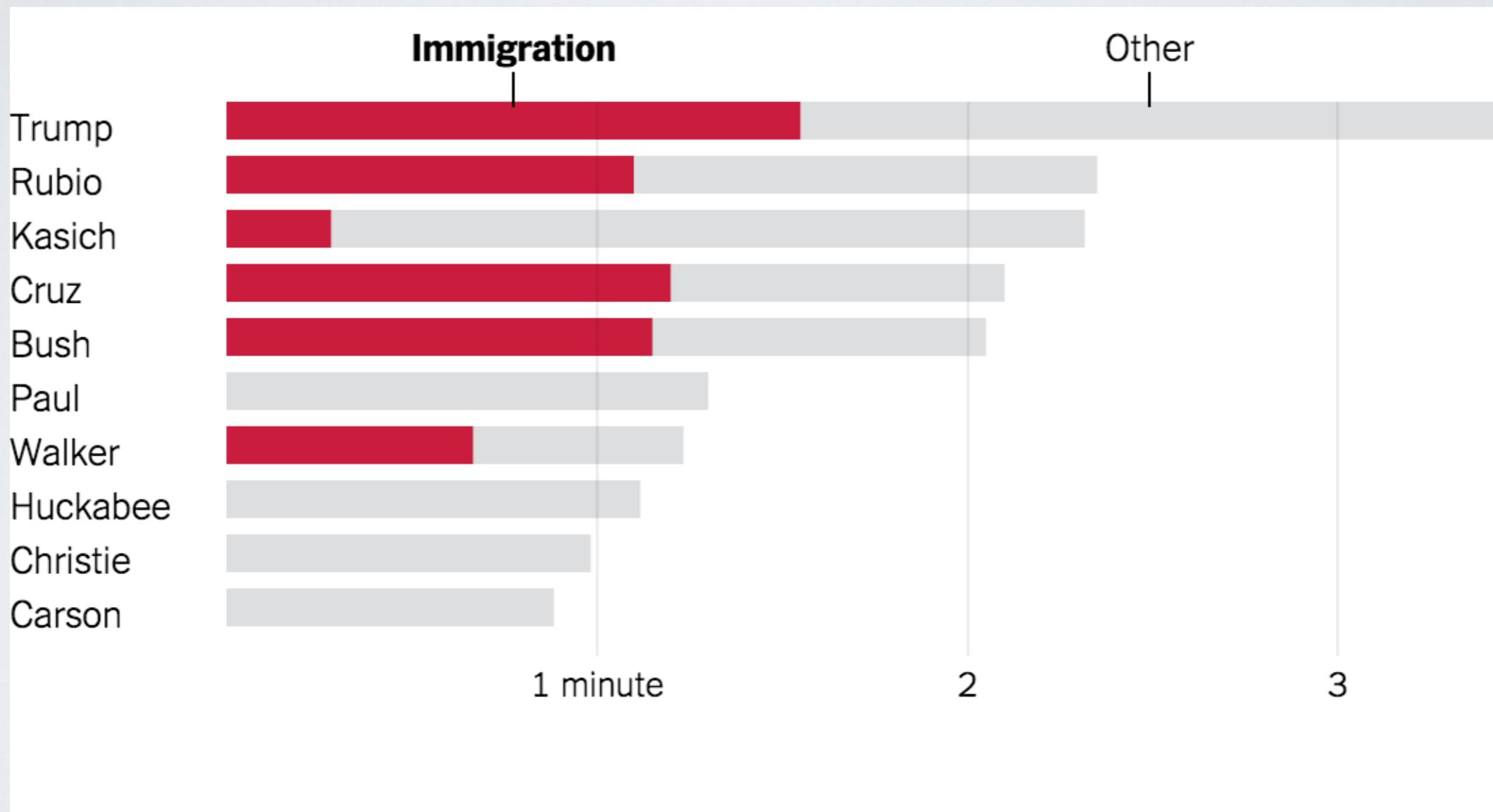
REPUBLICAN DEBATE: ANALYSIS AND HIGHLIGHTS



REPUBLICAN DEBATE: ANALYSIS AND HIGHLIGHTS



REPUBLICAN DEBATE: ANALYSIS AND HIGHLIGHTS



COLLABORATIVE RESEARCH

INTERDISCIPLINARY RESEARCH

- Transferability of Methods,
- Tools,
- Data,
- Concepts etc.

VIRTUAL RESEARCH ENVIRONMENTS

- VRE or Virtual Lab
- Supports collaboration of researchers across time and space
- Does not have to be interdisciplinary

CITIZEN SCIENCE

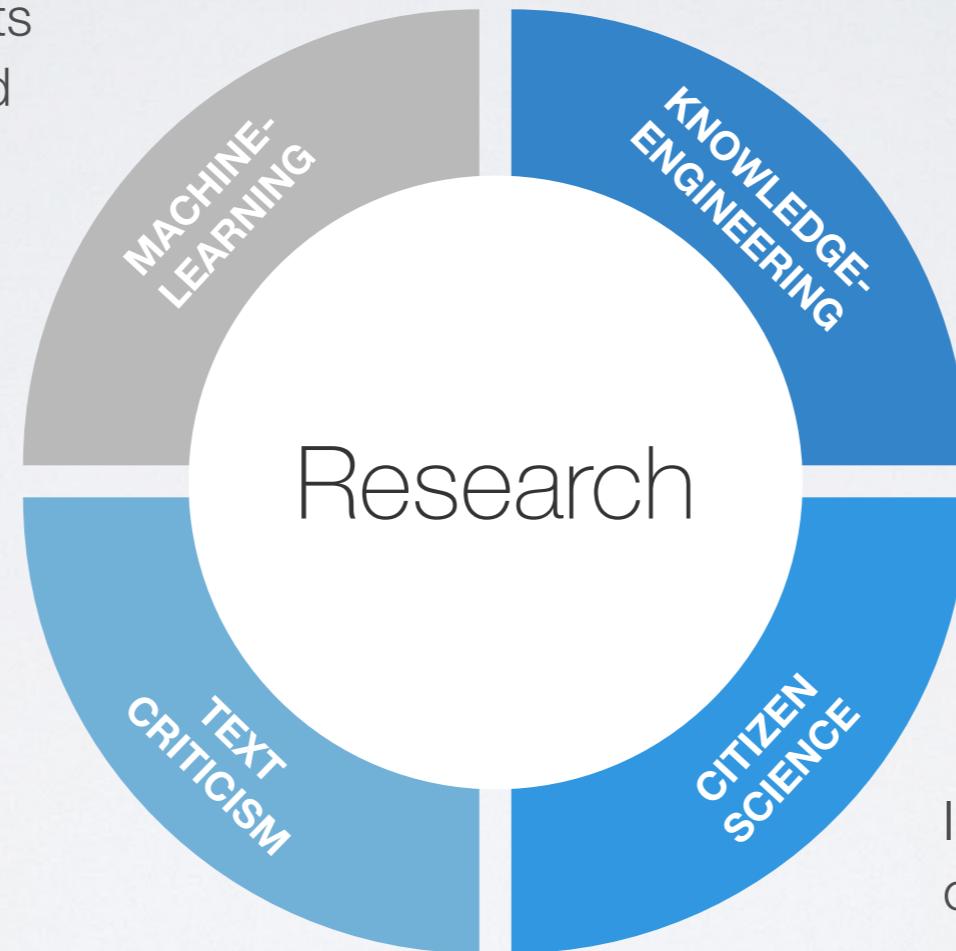
Thomas Köntges
Introduction to Digital Humanities
University of Leipzig



Digital Humanities
UNIVERSITÄT LEIPZIG

MY HEADACHE

Classification of manuscripts using machine-learning and human-computing-processes.



Editions in the age of open data and open access.
Lachmann/Mass in the digital age.

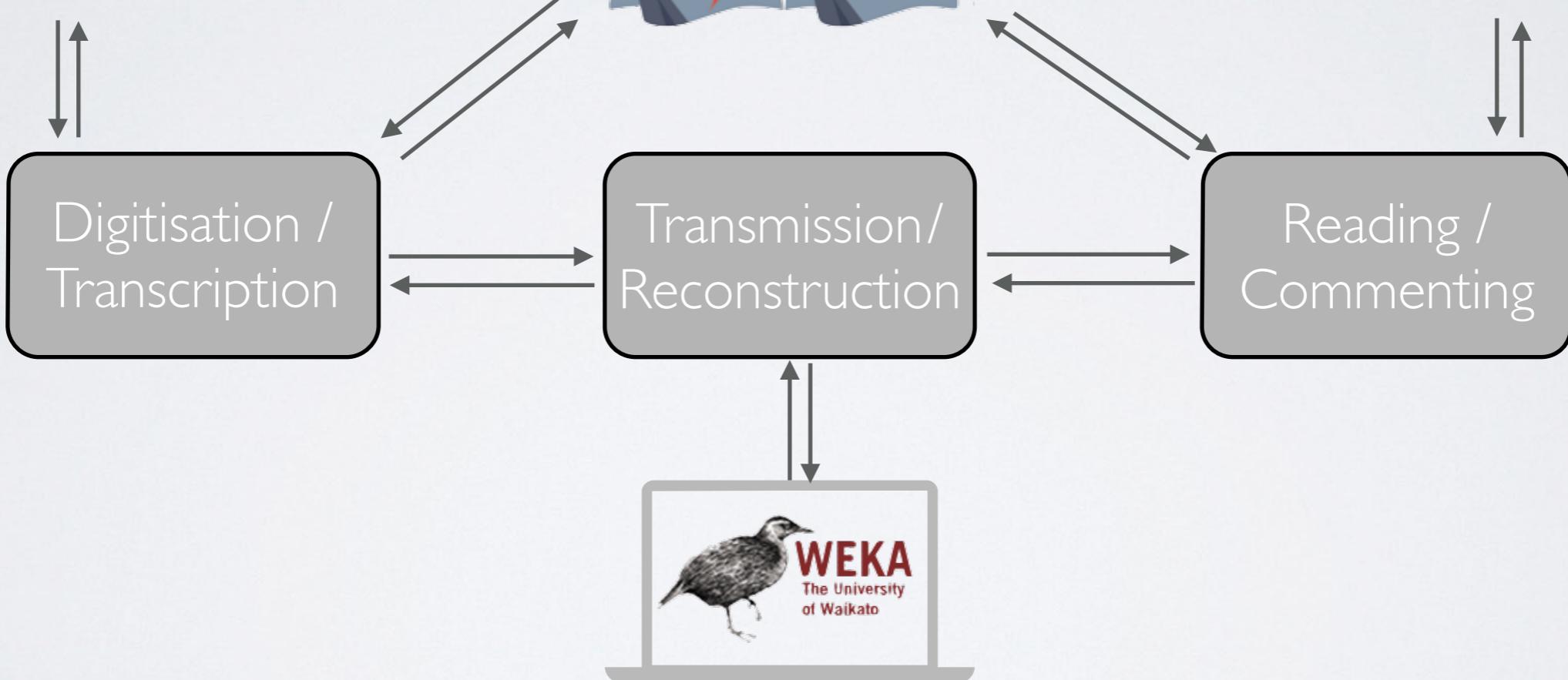
Computer-supported scoring of textual variants.
Scalability of textual data, image data, and learning environments.

In contrast to traditional crowdsourcing. Data quality management. Working with non-experts.

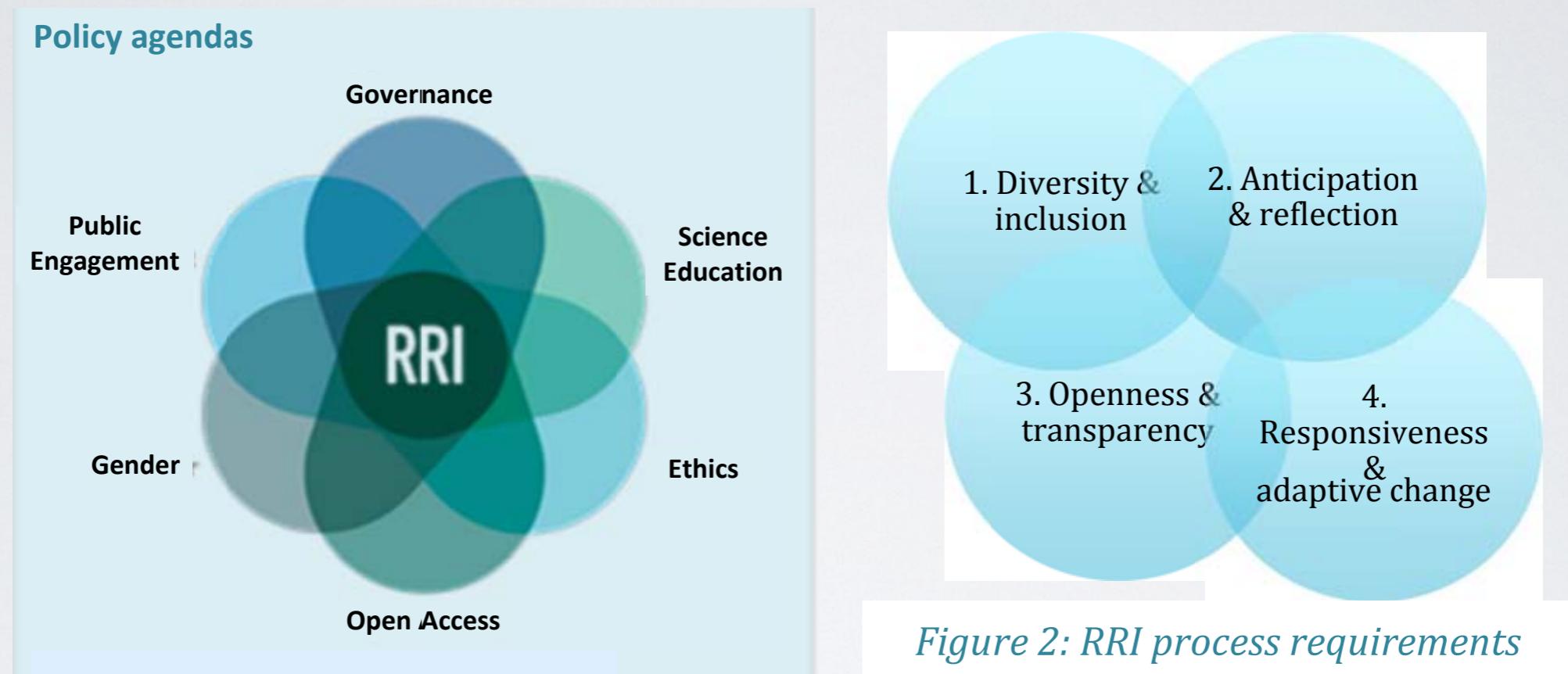
ARXE.TYPE WORKFLOW



Learners/Annotators

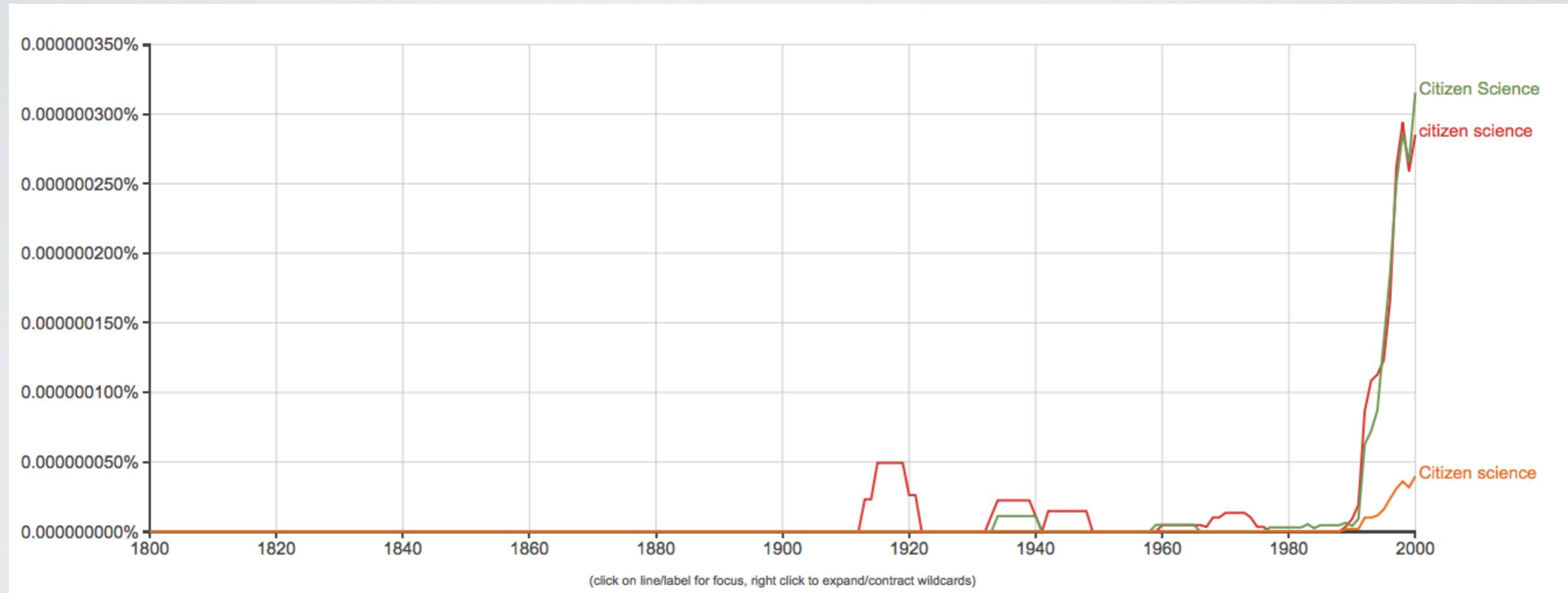


RRI



[http://www.rri-tools.eu/documents/10182/18424/
RRITools_D1.1-RRIPolicyBrief.pdf](http://www.rri-tools.eu/documents/10182/18424/RRITools_D1.1-RRIPolicyBrief.pdf)

A RECENT TREND?



WHO INVENTED IT?

- Wells Cooke's research on bird immigration
- Involvement of ornithologists in the Mississippi valley in 1881-1882
- Christmas Bird Count started in December 1900 by Frank Chapman in North America

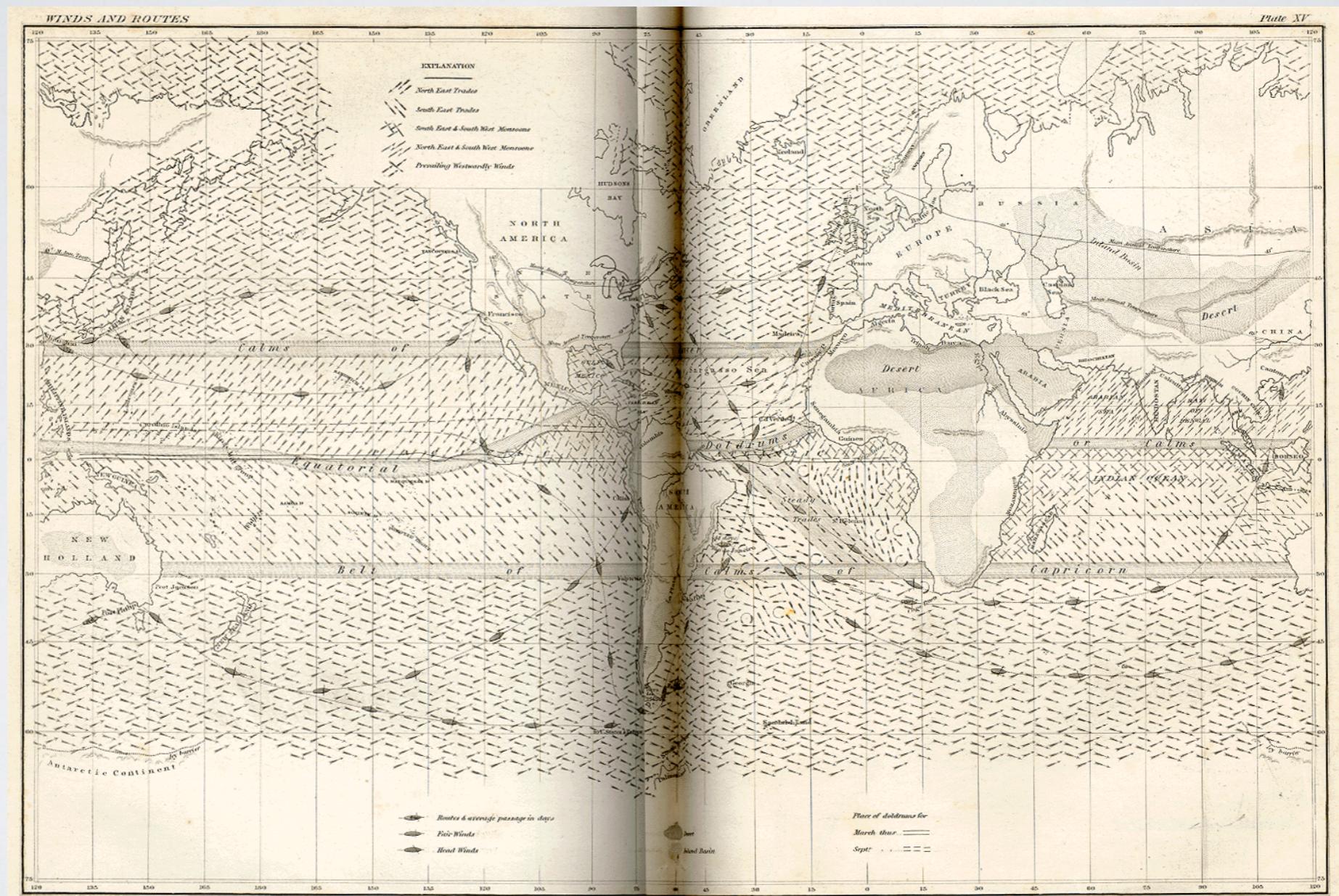
WHO INVENTED IT?

- American Ornithologists Union around 1883
- ‘Could you count the birds that strike the lighthouse?’
- ‘This is the dumbest thing I’ve ever heard of. Why am I counting birds?’

GLOBAL CITIZEN SCIENCE

- Citizens in 9 nations and colonies were sent to the coast to record tidal heights for the same two weeks in June 1835.
- William Whewell was awarded a Royal Medal
- In the late 1840s Matthew Maury asked farmers to telegraph weather observations in the US

WIND AND CURRENT CHART (1847)

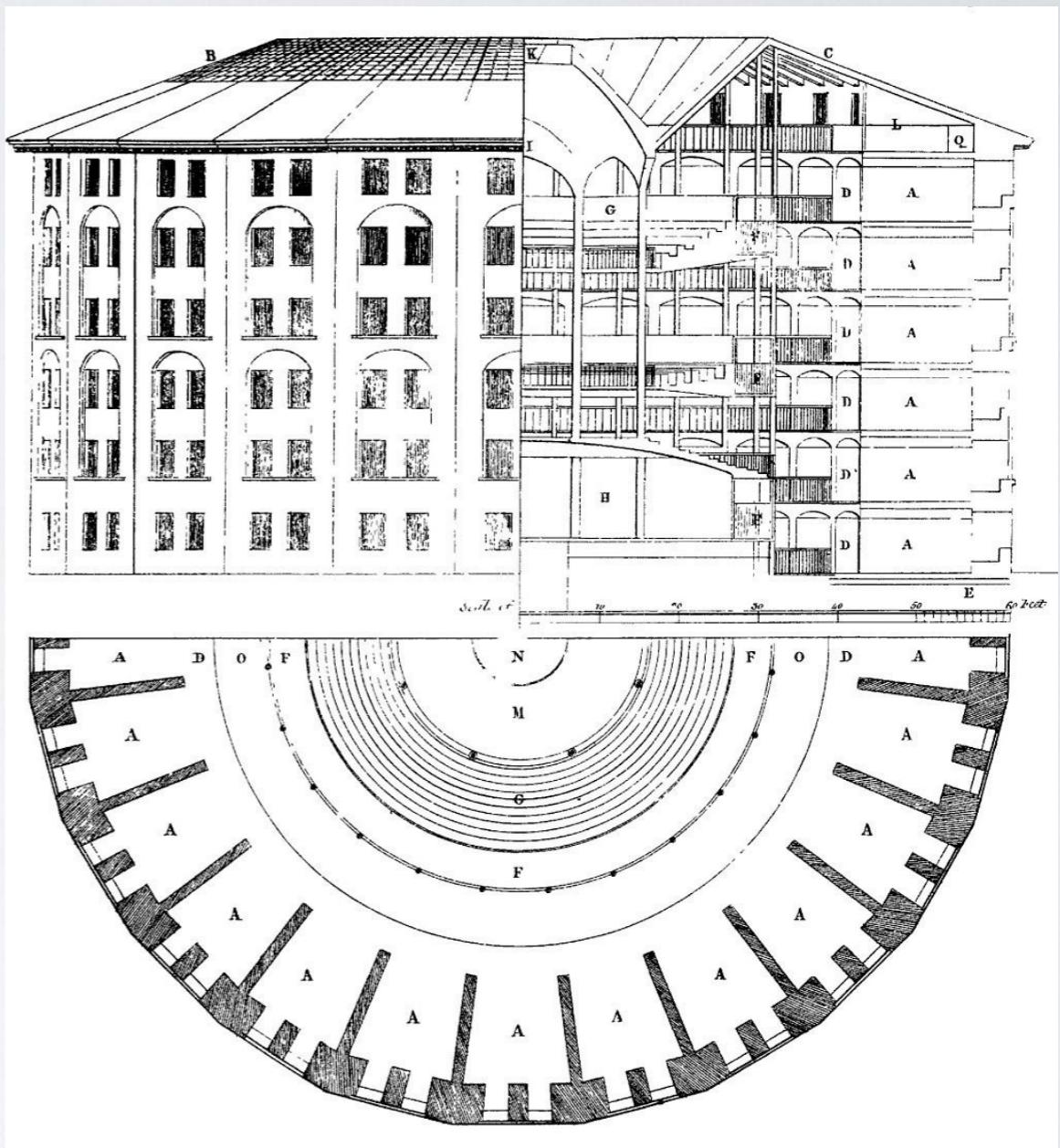


PHILOSOPHICAL APPROACHES

PANOPTICON VS. OLIGOPTICA

PANOPTICON

- Theory by Jeremy Bentham and Samuel Bentham 18th century
- Jeremy applies it to prison structure
- A few manager / guards can supervise many workers / prisoners



OLIGOPTICA

- Theory by Bruno Latour (2005)
- Applied by Chris Otter (2009) to Victorian Era (19th century)
- Victorian oligoptic visual economy with increased self-observation
- A good example of the Era is Self Help by Samuel Smiles

OLIGOPTICA

The greatest results in life are usually attained by simple means, and the exercise of ordinary qualities. The common life of every day, with its cares, necessities, and duties, affords ample opportunity for acquiring experience of the best kind [...]

Self Help, Samuel Smiles

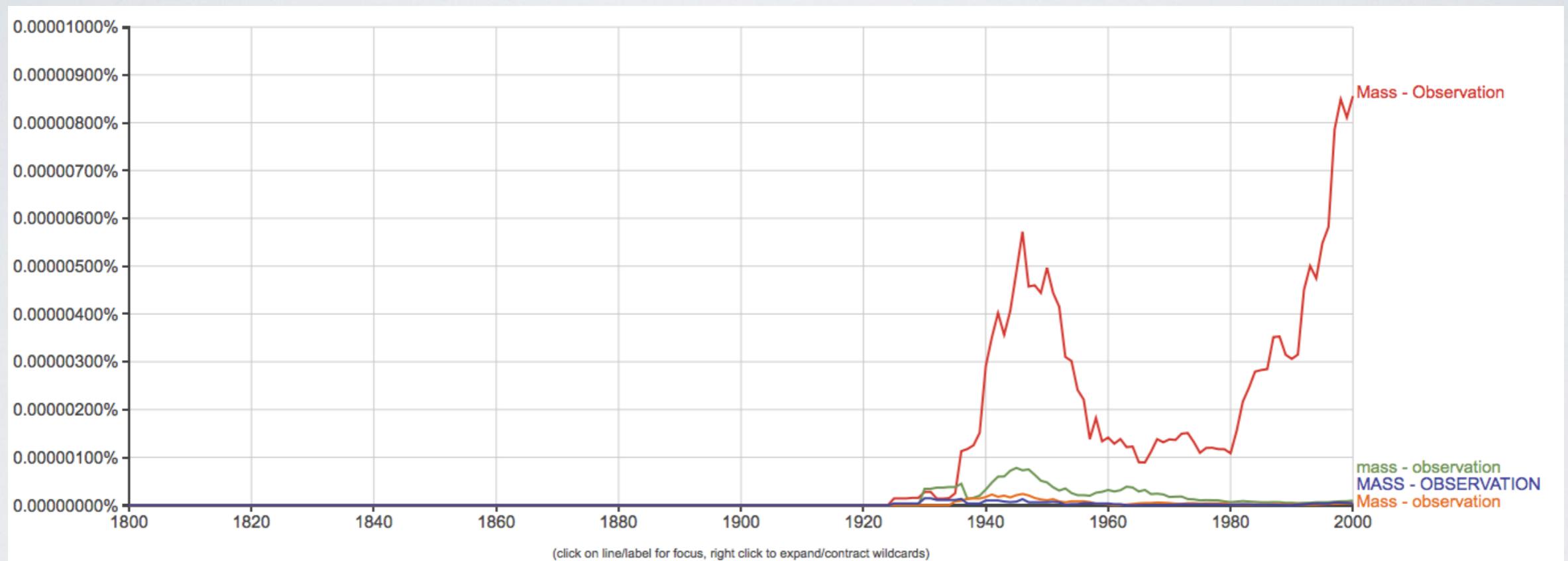
OLIGOPTICA

[...] Newton's whole mind had already been devoted for years to the laborious and patient investigation of the subject of gravitation; and the circumstance of the apple falling before his eyes was suddenly apprehended only as genius could apprehend it, and served to flash upon him the brilliant discovery then opening to his sight.

Self Help, Samuel Smiles

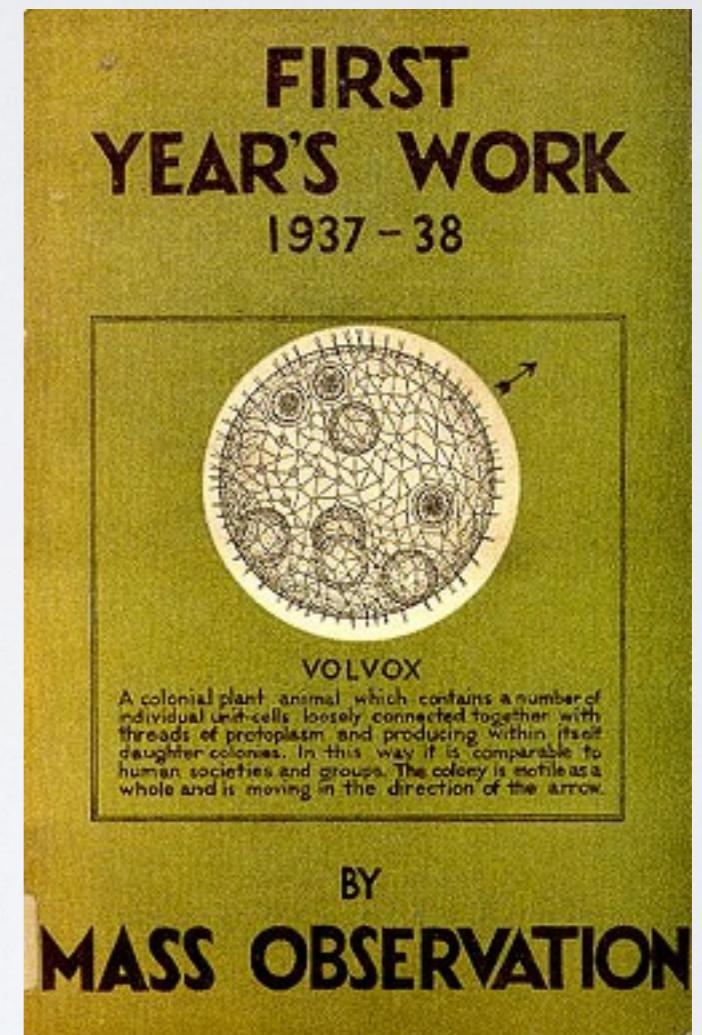
MASS OBSERVATION

MASS-OBSERVATION



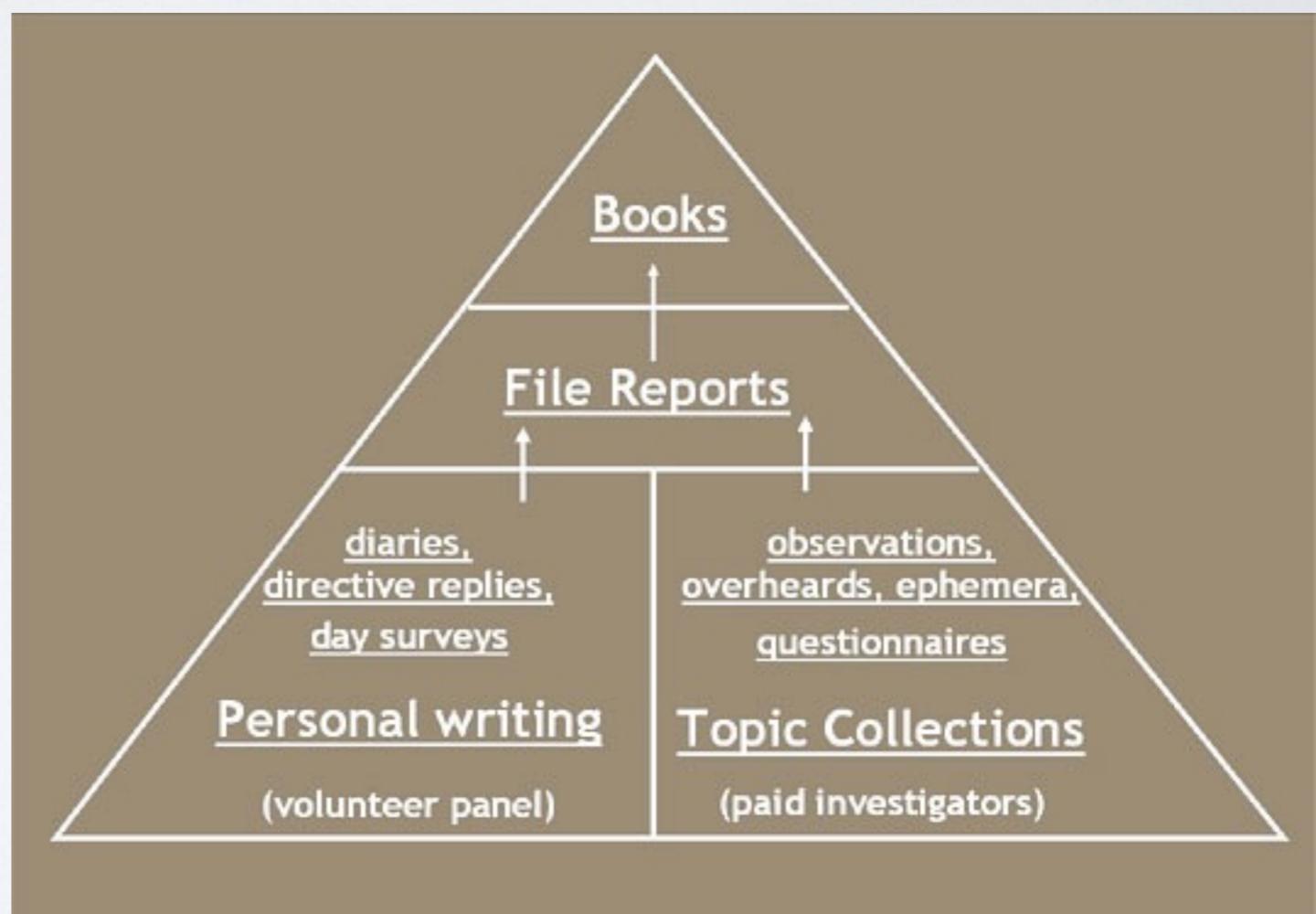
MASS-OBSERVATION

- Tom Harrison, Charles Madge, Humphrey Jennings
- Research Organisation founded in 1937
- 500 hundred untrained volunteers
- bidirectional process



MASS-OBSERVATION

- Worktown (Bolton, Greater Manchester)
- Metrop (A London borough)
- Also collecting emotions and feelings concerning disaster, hopes, and fears



MASS-OBSERVATION



Copyright Bolton Council

Photograph
by
Humphrey
Spender

(Painting
by William
Coldstream)

CITIZEN SCIENCE VS CROWDSOURCING

ADAM STEVENS

- “But (and you might have guessed there was a ‘but’ coming), let’s not pretend that it is something that it is not. It’s not, as far as I’m concerned, “citizen science”. **It is data crunching, plain and simple**, and I think it could be so much more. I’ve had a go at a number of the different projects and they were entertaining for about five minutes, after which I was often left thinking “yes, but what does it mean if that galaxy is elliptical and that one is spiral?” or similar.”

CROWDSOURCING

amazon mechanical turk
Artificial Artificial Intelligence

Your Account HITs Qualifications

Already have an account?
Sign in as a Worker | Requester

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

314,917 HITs available. [View them now.](#)

Make Money
by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → **Work** → **Earn money**

Find HITs Now

or learn more about being a Worker

Get Results
from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get Started.](#)

As a Mechanical Turk Requester you:

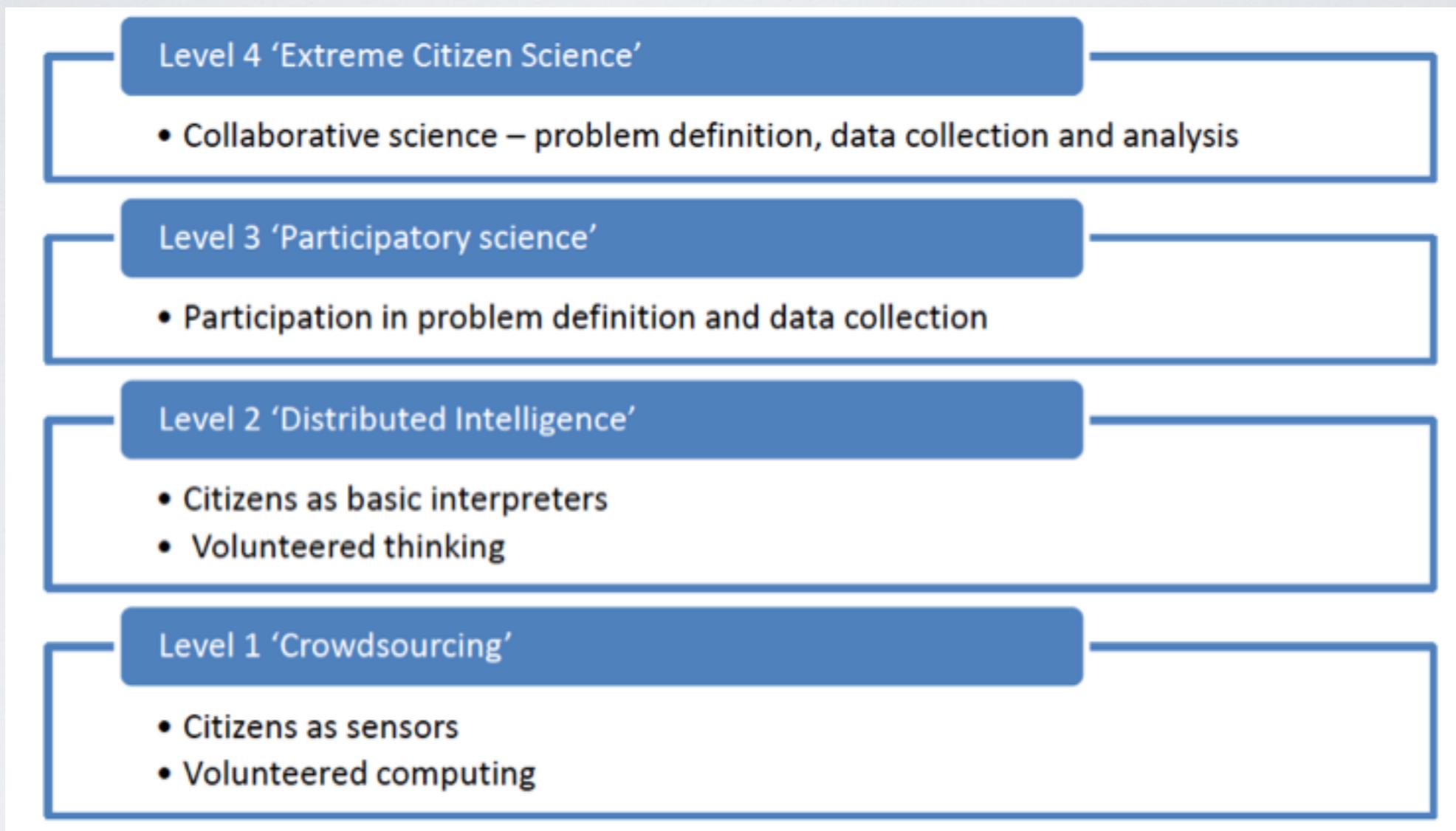
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your account → **Load your tasks** → **Get results**

Get Started

The screenshot shows the Amazon Mechanical Turk homepage. At the top, there are three tabs: 'Your Account' (highlighted in blue), 'HITs' (highlighted in yellow), and 'Qualifications'. To the right, there is a link 'Already have an account? Sign in as a Worker | Requester'. Below the tabs, there is a navigation bar with links to 'Introduction', 'Dashboard', 'Status', and 'Account Settings'. The main content area features a yellow gradient background. It starts with the text 'Mechanical Turk is a marketplace for work.' followed by 'We give businesses and developers access to an on-demand, scalable workforce.' and 'Workers select from thousands of tasks and work whenever it's convenient.' Below this, a red banner displays the text '314,917 HITs available.' followed by a blue link 'View them now.'. On the left side, there is a section titled 'Make Money by working on HITs' with a description of what HITs are and a link to 'Find HITs now.'. It also lists benefits for workers: working from home, choosing work hours, and getting paid for good work. Below this is a flow diagram with three circles: 'Find an interesting task' (with a blue arrow pointing to the next circle), 'Work' (with a blue gear icon inside), and 'Earn money' (with a dollar sign icon inside). A blue button at the bottom says 'Find HITs Now'. On the right side, there is a section titled 'Get Results from Mechanical Turk Workers' with a description of how requesters can use the service. It lists benefits for requesters: access to a global workforce, quick completion of tasks, and pay only for satisfaction. Below this is another flow diagram with three orange circles: 'Fund your account' (with a plus sign icon inside), 'Load your tasks' (with a document icon inside), and 'Get results' (with a star icon inside). An orange button at the bottom says 'Get Started'.

M. HAKLAY'S LEVEL OF PARTICIPATION



INVOLVEMENT IN SCIENCE TYPOLOGIES

Stage of Inquiry	Cooper et al.	Wilderma	Bonney et al.	Contributory	Collaborative	Co-created
Define question	✓	✓	✓			X
Gather information			✓			X
Develop hypotheses			✓			X
Design study	✓	✓	✓		(X)	X
Data collection	✓	✓	✓	X	X	X
Analyze samples		✓	✓		X	X
Analyze data	✓		✓	(X)	X	X
Interpret data	✓	✓	✓		(X)	X
Draw conclusions	✓		✓		(X)	X
Disseminate results			✓	(X)	(X)	X
Discuss results & ask new questions			✓			X

TABLE I
VOLUNTEER INVOLVEMENT IN ENVIRONMENTAL SCIENCE TYPOLOGIES,
WITH DEFINITIONS OF PARTICIPATORY SCIENCE MODELS. ✓ = INCLUDED
IN MODEL; X = PUBLIC INCLUDED; (X) = PUBLIC SOMETIMES INCLUDED.

THE ABOVE
STILL FOCUSED ON THE
SCIENCE INSTEAD OF THE
SCIENTIST

CITIZEN SCIENCE VS. COMMUNITY SCIENCE

TOM O'REILLY (CREATE MORE VALUE THAN YOU CAPTURE, 15 JUNE 2009)

“It's about how you can add value to the communities that happen to include you [...] Not surprisingly the more value you create for your community, the more value they will create for you.”

TYPES OF CITIZEN SCIENCE

- Action
- Conservation
- Investigation
- Virtual
- Education

ACTION

- Example: Sherman's Creek Conservation Association
- Issues:
 - Challenges for the aggregation of data
 - Organisation seldom scales
 - Minimal use of technologies

CONSERVATION

- Example: Northeast Phenology Monitoring Project
- Issues:
 - Focus on generating data primarily for decision-making
 - Often not in easily accessible format
 - Heavily dependent on state funding

INVESTIGATION

- Example: The Great Sunflower Project
- Issues:
 - Valid scientific results
 - Almost never utilise volunteer screening
 - Sampling bias

VIRTUAL

- Example: Galaxy Zoo
- Issues:
 - Ensuring valid scientific results
 - Critical mass of contributors
 - Top-down organisations by academics
 - Complex custom web-platforms

ACTION

- Example: Fossil Finders
- Issues:
 - Citizen science often only because it has a formal research partner
 - Relative costs of acquiring data via formal learning environments are high
 - Valid results matter less than the learning itself

ADVANTAGES

- Sparking Interest and Excitement
- Understanding Scientific Content and Knowledge
- Engaging in Scientific Reasoning
- Reflecting on Science
- Using the Tools and Language of Science
- Identifying with the Scientific Enterprise

CITIZEN SCIENCE

Citizen Science represents a **new type of open movement** welcoming **contributions** to **scientific research** from a **diverse** population of volunteers.

DIGITAL HUMANITIES: IT'S YOUR TURN

- What does the field attempt to answer?
- How would research data look like in this particular field? Give examples.
- How would the research data be structured? Which digital format can express the structure?
- What are the computational challenges in the field? What can computational results express? What are potential limitations?

QUESTIONS?

