Accelerating compute-intensive humanities through collaborative development

OR 'when a few data points are not enough'

Kristoffer L Nielbo

kln@cas.dk

knielbo.github.io

chcaa.io

Center for Humanities Computing
Aarhus University, Denmark

Outline

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

On the role of informatics

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

the data deluge is transforming knowledge discovery and understanding in every domain of human inquiry

a large part of these data are soft and unstructured $\Rightarrow$ to get value from these data, humanities (and social sciences) must utilize automation

human informatics - automatic information processing in the humanities
$\Rightarrow$ *use of automation in the study of culture and history*

K. Steinbuch (1957) Informatik: Automatische Informationsverarbeitung [Berlin: SEG-Nachrichten]

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Data, automation &
humanities
On the role of informatics
Antithesis to informatics
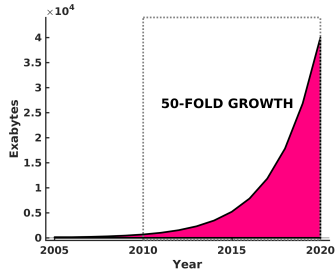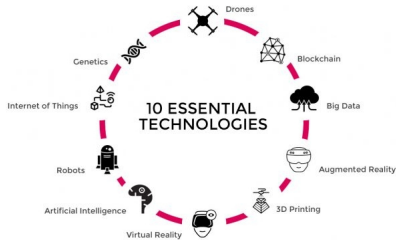Learning curves
Sine qua non
CHCAA
Modes of intervention
Project life-cycle
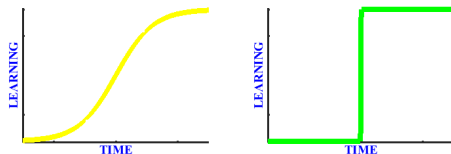Applications of
automation
Sanity checks
Old problems
New knowledge

humanities and arts are the anti-thesis' of informatics ⇒ research that solely relies on very few data points, an ideological resistance toward formalization, and a "myopic" perspective and human computation

enable and empower domain knowledge in history, language, literature &c combined with microscopic and (predominantly) qualitative analysis of human cultural manifestations

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

'digital tools' often behaves as the the research equivalent of 'plug and pray'

"breaking through the informatics learning curve" only works if humanities researchers do what they are good at: *humanities research*

Formalization is necessary

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Data, automation &
humanities
On the role of informatics
Antithesis to informatics
Learning curves
Sine qua non

CHCAA
Modes of intervention
Project life-cycle

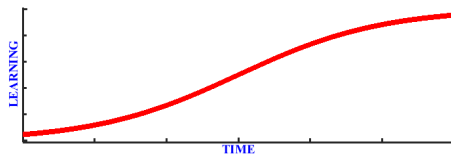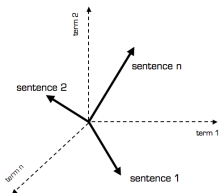Applications of
automation
Sanity checks
Old problems
New knowledge

> One by one, the snowflakes floated down on to his warm snout, and melted. He reached out to grab them so he could admire them for a fleeting moment. He looked towards the sky and watched them drift down towards him, more and more, soft and light as a feather. "So that's how it works," thought Moomintroll. "And I thought somehow that the snow grew from the ground up
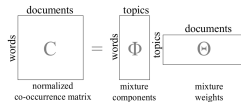
— Tove Jansson, Moominland Midwinter

an understanding of and willingness to *describe cultural/historical systems and processes in terms of explicit rules and procedures* are more important than practical programming skills.

co-creation of explicit formalism for a research problem is the key to successful collaboration

Modes of intervention

Accelerating compute-intensive humanities through collaborative development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Data, automation & humanities
On the role of informatics
Antithesis to informatics
Learning curves
Sine qua non

CHCAA
Modes of intervention
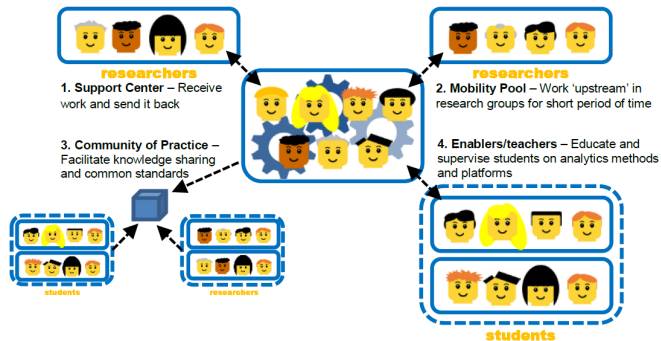Project life-cycle

Applications of automation
Sanity checks
Old problems
New knowledge

researchers
1. Support Center – Receive work and send it back

researchers
2. Mobility Pool – Work 'upstream' in research groups for short period of time

3. Community of Practice – Facilitate knowledge sharing and common standards

4. Enablers/teachers – Educate and supervise students on analytics methods and platforms

students

researchers

students

In order to enable and empower domain knowledge in the humanities, AU.ARTS established Center for Humanities Computing Aarhus in 2019.

Project life-cycle

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Project can have any size or shape, e.g., simple web page or data hosting service to large-scale AI applications.

✳ `Project preparation` (submit ticket, build business case, identify stakeholders, project charter)

✳ `Project start-up` (solution architecture + dev/delivery plan, prioritize requirements, project abstract, communication plan)

✳ `Development` [ITERATE] (timeboxing with delivery on 1-2 week sprint cadence, hack-days)

✳ `Deployment` [ITERATE] (assemble, review and deploy)

✳ `Project finalization` (assessment and transference of rights/project to project owner)

Benchmarks & sanity checks

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Data, automation &
humanities
On the role of informatics
Antithesis to informatics
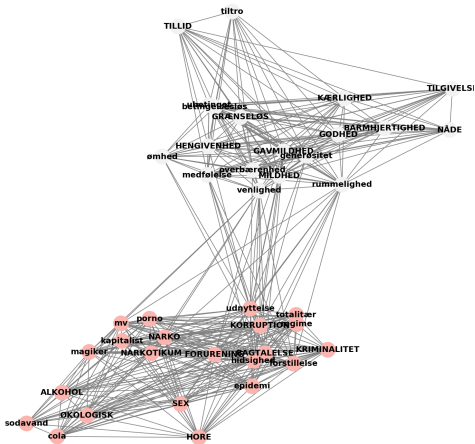Learning curves
Sine qua non

CHCAA
Modes of intervention
Project life-cycle

Applications of
automation
Sanity checks
Old problems
New knowledge

In $\sim$ 12K sermons from the Evangelical Lutheran Church in Denmark (2011-16), we find a associative system that reflects a traditionalist value system confronted with contemporary issues.

Solving old problems

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Figure 1: Debate regarding the bipartite composition *Gesta Danorum*, 1) is the transition between the old mythical and new historical parts located in book eight, nine, or ten? and 2) is this transition gradual or sudden?

Nielbo, K., M.L. Perner, C. Larsen, J. Nielsen, D. Laursen (2019) Automated Compositional Change Detection in Saxo Grammaticus' Gesta Danorum. (hal-02084682)

Discovery of new knowledge

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Data, automation &
humanities
On the role of informatics
Antithesis to informatics
Learning curves
Sine qua non

CHCAA
Modes of intervention
Project life-cycle

Applications of
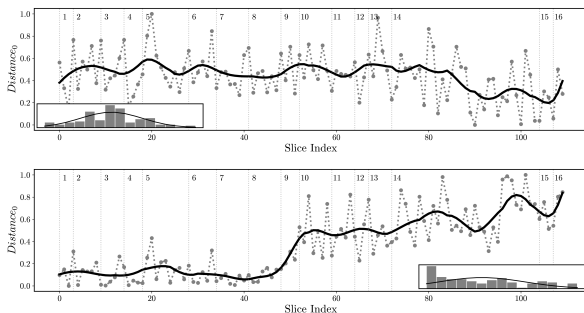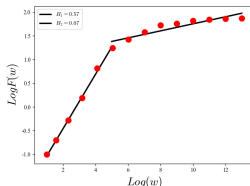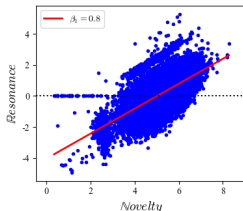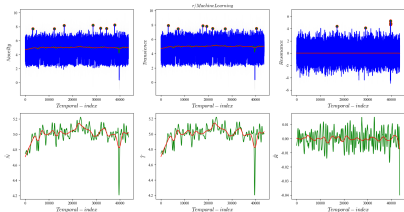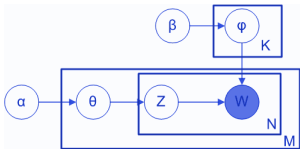automation
Sanity checks
Old problems
New knowledge

Trend detection for social media works well for point-like events (e.g., natural catastrophes and epidemics), but what are the signature(-s) of social trends? Analyzing 7TB+ data from reddit.com, we find that in certain subreddits, novelty ressonates more with the future and that content display long-term memory at short and intermediary time-scales.

Accelerating
compute-intensive
humanities through
collaborative
development

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Data, automation &
humanities
On the role of informatics
Antithesis to informatics
Learning curves
Sine qua non
CHCAA
Modes of intervention
Project life-cycle
Applications of
automation
Sanity checks
Old problems
New knowledge

Thank you for your attention

kln@au.dk
knielbo.github.io
chcaa.io

slides: http://knielbo.github.io/files/kln_itvest19.pdf