research technology and scale

the digital curriculum 2020-21

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

Center for Humanities Computin Aarhus|chcaa.io
Aarhus University, Denmark

August 13, 2020

Outline

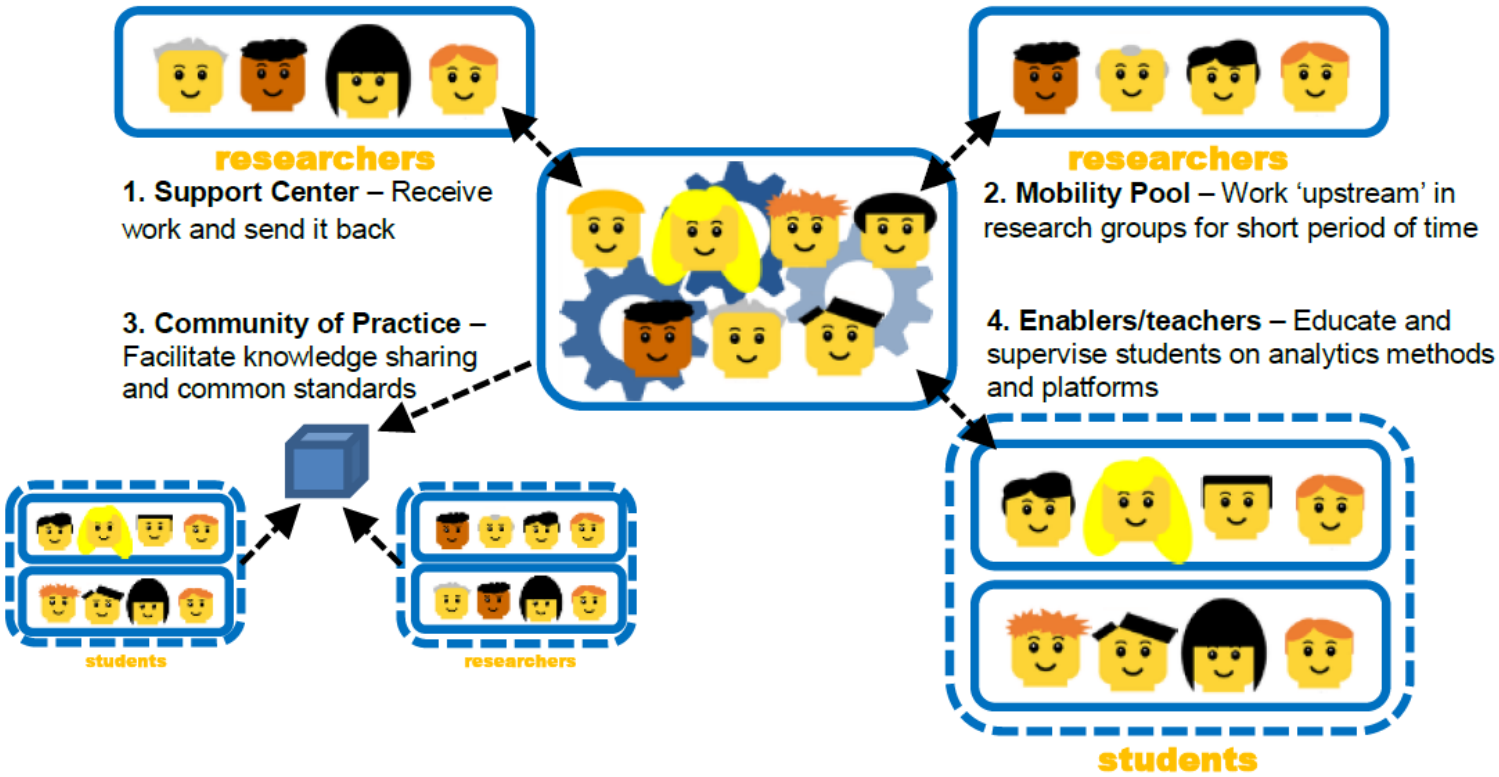
research it supports research with a coordinated set of services across a range of computation and data analysis needs

IS a substantial part of eScience

NOT it department (infrastructure and maintenance of it systems)

NOT educational it (facilitate learning through it solutions)

research it solutions tend to bleed into education as a bottom-up process



research it at the faculty of arts, aarhus university

together with rit personnel from other dk universities, currently supports
DIGITAL LITERACY$_{research}$ & *DIGITAL CURRICULUM*$_{education}$

"digital" is a methods-issue in relation to research & education

CENTER FOR HUMANITIES COMPUTING AARHUS

"digital methods" $\sim$ a question of SCALABILITY in response to digitization

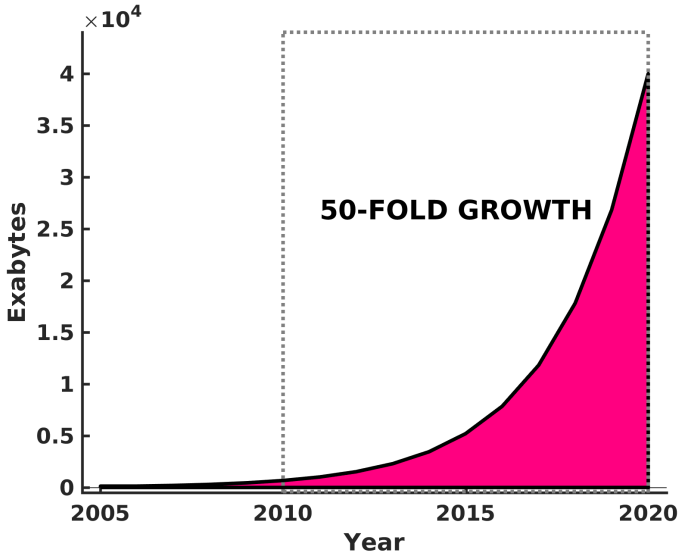
assumption :: scalability ... fundamentally changes how we do research

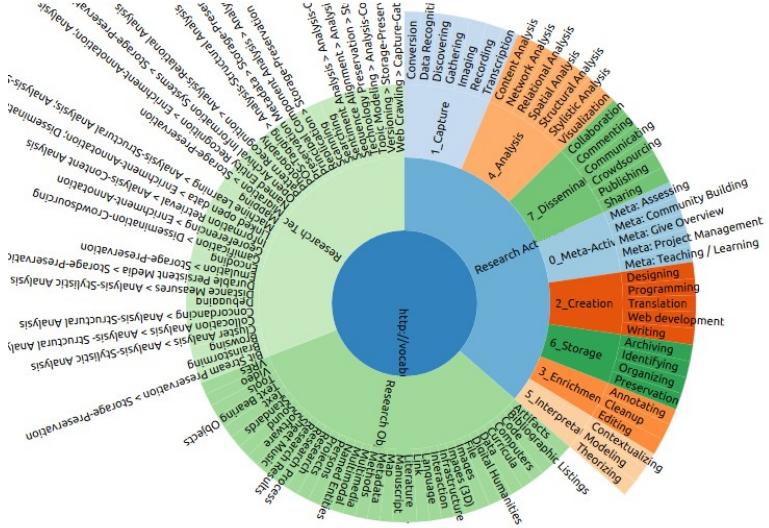
corollary :: scalability ... requires algorithmic automation and algorithms depend on models of structures and processes

$\Rightarrow$ "digital methods" are not a methods-issue $\sim$ TECHNOLOGY

a) scale in research technology and b) scaling research technology for education

initially, 'scaling ... for education' seemed like the simple task





data management at scale

"as in other research domains, data became the promised land for humanities and arts ... three years ago a data set was measured in mbs, now 20-30gb is the standard and we are seeing many data sets on the tb-scale."

data :: **large, soft & heterogeneous**
dual problem :: data are **sensitive & restricted** access
derived problems::

- data are relatively easy to access
- no standardized procedures for risk-benefit evaluation
- research evolves at a faster scale than legal
- diversification of tools for management and analysis

data-related problems multiplies in education

proper research practice requires **data management principles**

Findable Accessible Interoperable Re-usable

for all data sets:

- F::PID & rich metadata are indexed
- A::standardized and open protocol for retrieval & ++persistent metadata
- I::use controlled vocabulary for metadata & references
- R::described w. license, provenance & domain-relevant standards

⇒ one of two ways, revive W3C's semantic web or 'just' **implement common sense for research data**

teach operational standards for large, soft, and heterogeneous data

proper tooling makes researchers and educators' lives *a lot* easier

$\Rightarrow$ demand for flexible tools is a response to tool diversification and large, heterogeneous data

flexibility/usability tradeoff :: flexible tools satisfy more requirements $\rightarrow$ complexity$\uparrow$ & usability$\downarrow$

SSH have historically solved the f/u tradeoff with a gui-based model

gui or cli solutions for research and education

graphical user interface ::

- visual approach to computer interaction
- fast learning curve & **usability**↑
- 'plug-n-play' solution with limited flexibility

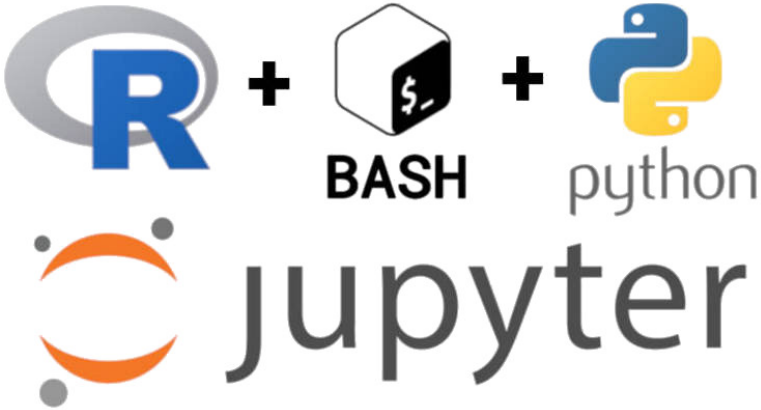
command line interface ::

- text-based approach to computer interaction
- resource efficient & **flexibility**↑
- expert-friendly solution with limited usability

interactive solutions (flexibility↑ & usability↑) are gaining traction in research

interactive computing in a dual sense

- tools that allows users to enter commands and data interactively
- tools that users develop and run collaboratively



jupyter :: interactive environment that combines visual and text-based approaches with storytelling

goal to offer cloud-based interactive computing to researchers and students

cloud-based it-infrastructure provides on-demand services and resources via the internet

- universities provide access to commercial cloud vendors
- deic will provide access to interactive computing in the cloud

summary

- research-it support for project development
- support use of research technology in class
- scaling to *n* students is *the challenge*
- proper data management practice should always be included
- prioritize solutions that utilize interactive computing
- work towards a cloud-based solution

THANKS

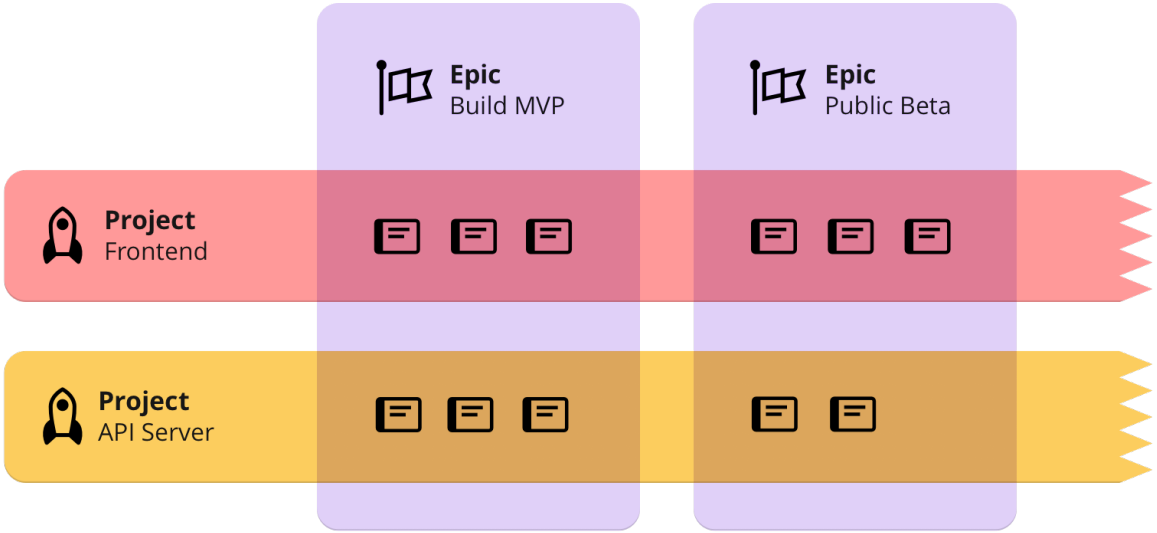
kln@au.dk
knielbo.github.io
chcaa.io

SLIDES

knielbo.github.io/files/kln_tscale.pdf

OR

CENTER FOR HUMANITIES COMPUTING AARHUS



Managing multiple DC (sub-)projects

DIGITAL CURRICULUM projects are thin (composed of relatively few epics & stories) and fuzzy (projects bleed into each other)

- maximize tool re-use within epics (project phases) → invest in flexible tools
- share compute resources between projects → project collaboration
- accept functional divisions → construct a common vocabulary

and, proper TOOLING will make your activities a lot easier