

workflow, technologies, and rules of engagement

digital literacy 2020-21

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

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



August 25-26, 2020

Outline

- 1 research it
 - clarification
 - center for humanities computing
- 2 workflow & ROE
 - "digital methods"
 - ROE
 - workflow
- 3 data management
 - dm at scale
 - standards in dm
- 4 tools
 - proper tooling
 - gui or cli
 - interactive computing
- 5 cloud computing
- 6 summary

research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary



CENTER FOR HUMANITIES
COMPUTING AARHUS



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 and development for scientific inquiry and knowledge discovery

research it

clarification

center for humanities
computing

workflow & ROE

"digital methods"

ROE

workflow

data management

dm at scale

standards in dm

tools

proper tooling

gui or cli

interactive computing

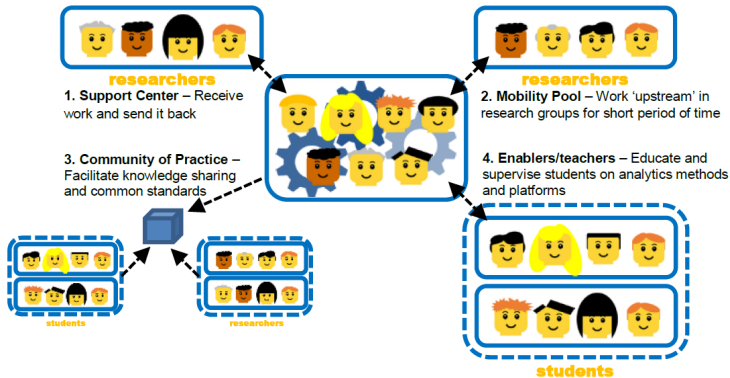
cloud computing

summary



CENTER FOR HUMANITIES
COMPUTING AARHUS





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 conceptualized as a methods-issue in relation to research



CENTER FOR HUMANITIES
COMPUTING AARHUS



research it
 clarification
 center for humanities
 computing
 workflow & ROE
 “digital methods”
 ROE
 workflow
 data management
 dm at scale
 standards in dm
 tools
 proper tooling
 gui or cli
 interactive computing
 cloud computing
 summary

"digital methods" ~ 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

⇒ "digital methods" are not a methods-issue ~ TECHNOLOGY

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

'scale in research technology is the bread and butter of research it

rules of engagement ~ initialization of the support team

- WHAT IS YOUR RESEARCH QUESTION
- WHERE ARE THE DATA THAT CAN SOLVE THAT RQ

research it
clarification
center for humanities
computing

workflow & ROE

"digital methods"

ROE

workflow

data management

dm at scale

standards in dm

tools

proper tooling

gui or cli

interactive computing

cloud computing

summary



CENTER FOR HUMANITIES
COMPUTING AARHUS



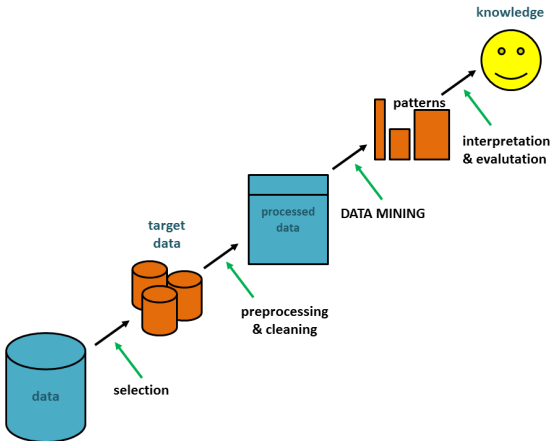


Figure: sequence of dependent phases (**epics**) that every research project travels through, e.g., data selection, data analysis. each epic is a collection of research tasks (**stories**) , e.g., extract data, train model. together stories and epics have the goal of **creating new knowledge**.

research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary

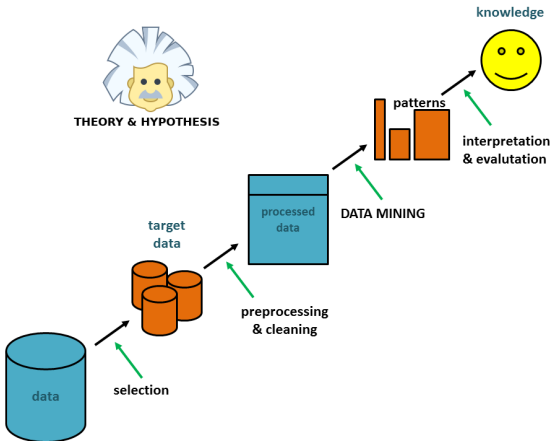
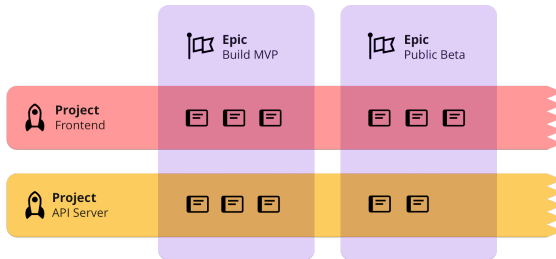


Figure: sequence of dependent phases (**epics**) that every research project travels through, e.g., data selection, data analysis. each epic is a collection of research tasks (**stories**) , e.g., extract data, train model. together stories and epics have the goal of **creating new knowledge**.

research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary

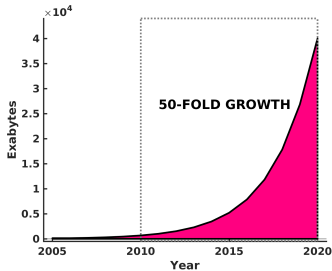


Managing multiple DL projects

DIGITAL LITERACY 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

research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary



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."

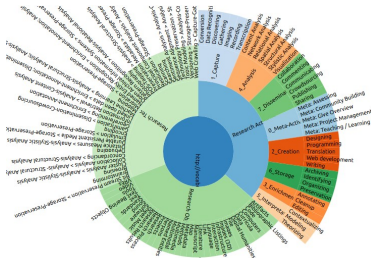
data :: large, soft & heterogeneous

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

operational data management standards are a must



research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary

proper research practice requires **data management principles**

Findable **A**ccessible **I**nteroperable **R**e-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**

DL will experiment with operational standards of FAIR

research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary



CENTER FOR HUMANITIES
COMPUTING AARHUS



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

⇒ demand for flexible tools is a response to tool diversification and large, heterogeneous data

flexibility/usability tradeoff :: flexible tools satisfy more requirements → complexity↑ & usability↓

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

research it
clarification
center for humanities
computing
workflow & ROE
“digital methods”
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary



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



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

research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary



CENTER FOR HUMANITIES
COMPUTING AARHUS



summary

- research-it support for project development
- support use of research technology
- scaling can (and most likely will) change your research
- proper data management practices should always be included
- prioritize interactive computing
- work towards a cloud-based solution

research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary



THANKS

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

SLIDES

knielbo.github.io/files/kln_dlopen.pdf

OR



research it
clarification
center for humanities
computing
workflow & ROE
"digital methods"
ROE
workflow
data management
dm at scale
standards in dm
tools
proper tooling
gui or cli
interactive computing
cloud computing
summary



CENTER FOR HUMANITIES
COMPUTING AARHUS

