Humanities & Arts Data Service
Every Researcher's Dream OR an Administrative Nightmare

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

Center for Humanities Computing Aarhus|chcaa.io
Aarhus University, Denmark
"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 & unstructured  
**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  

"center for humanities computing was a centralized response to the growing impact of data on humanities and arts"
OS requirements

three principles of open of the open science (OS) movement

**OPEN DATA::** FAIR principles implemented with the aid of a data steward and supervised by legal

**OPEN SOURCE::** interactive notebooks hosted at some (read: github) open repository
OS solutions

W3’s DCAT RDF vocabulary

- **FAIR::** common sense for research
- **CHALLENGE::** semantic web meets large, soft and unstructured data
- **CHALLENGE::** copyright meets open data
- **SOLUTION::** DCAT for metadata with quick-and-dirty risk evaluation + legal hotline

- **OSS::** common sense for research
- **CHALLENGE::** gui tools are almost always commercial closed source
- **CHALLENGE::** notebooks promote opaqueness and scale terribly
- **SOLUTION::** notebooks for interactive use only & reverse engineer + re-implement gui functionality
THANKS

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

SLIDES


OR