15-30 minutes of

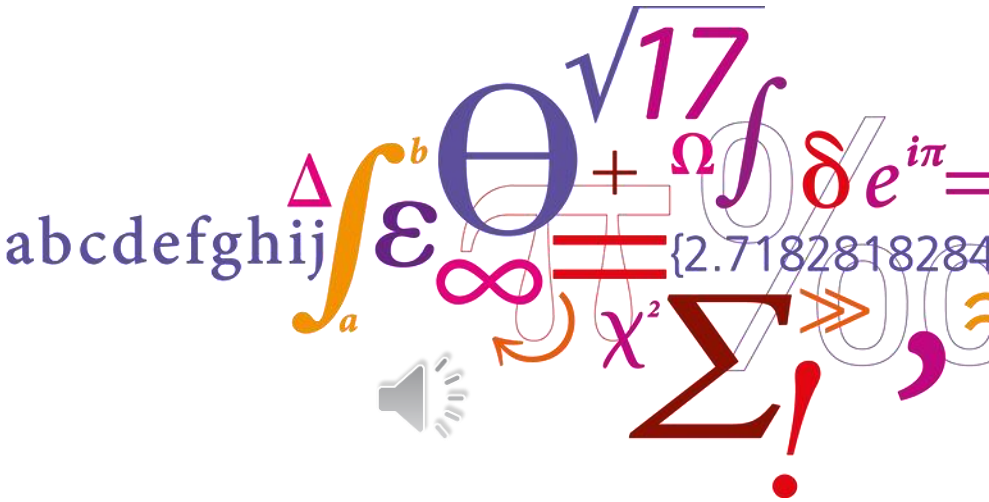
Data & Data Science Tools

Videndeling, inspiration, hænder-på

Koordineret og eksekveret af
DTU Bibliotek

V Jeannette Ekstrøm
Senior Information Officer
jeek@dtu.dk

Twitter: @JEkstroem



DTU Library
Technical Information Center of Denmark

Data kompetenceudvikling og ”awareness”

- DTU Smart Library
- DTU Code Club*
- Open Access
- Open Data ~ RDM
- Metadata & søgning
- TDM & licenser
- Data Visualisering
- Coding & scripting
- Interne processer
- ”**V**(i e)**R** data”

Ideen inspireret fra CreatingKnowledge 2018

https://noril.uib.no/article/view/2764

15 minutes of Data på Uppsala Universitet

Put a frame on it.

Anchor it.

Motivate the group.

Involve.

Activate.

Follow a path.

Plan.

Formålet med ”konceptet”

Hvem kan have glæde af dette… ?

- **Kolleger fra ALLE TEAMS er velkomne**

- **Hvorfor**
  - Fordi vi alle kan blive klogere sammen
  - Fordi formidling af,- metoder til, adgang til, licenser for DATA er relevant for os alle - fra systemfolk til undervisertyper

- (**META)Læringsmål for samtlige forløb** er, at vi alle får en ide om DATA – både i relation til adgang, håndtering, trends, formidling og politikker – og at vi føler os bedre rustet til vores opgaver.

  Desuden er forhåbningen, at alle kan få lidt inspiration og viden om, hvor vi selv skal sætte ind omkring ”kompetenceudvikling”, hvad kunne passe til vores faglige opgaver og faglige interesser.

  DATA er i alt, både job og privat……

https://luminariaz.files.wordpress.com/2017/11/the-fourth-industrial-revolution-2016-21.pdf

Fokus vil være, at gøre alle ”data literate”

- **Data literacy** is the ability to read, understand, create and communicate data as information.

  Much like literacy as a general concept, data literacy focuses on the competencies involved in working with data. As data collection and sharing become routine and data analysis and big data become common ideas in the news, business,[1] government[2] and society[3], it becomes more and more important for students, citizens, and readers to have some data literacy.

  » Source: Wikipedia

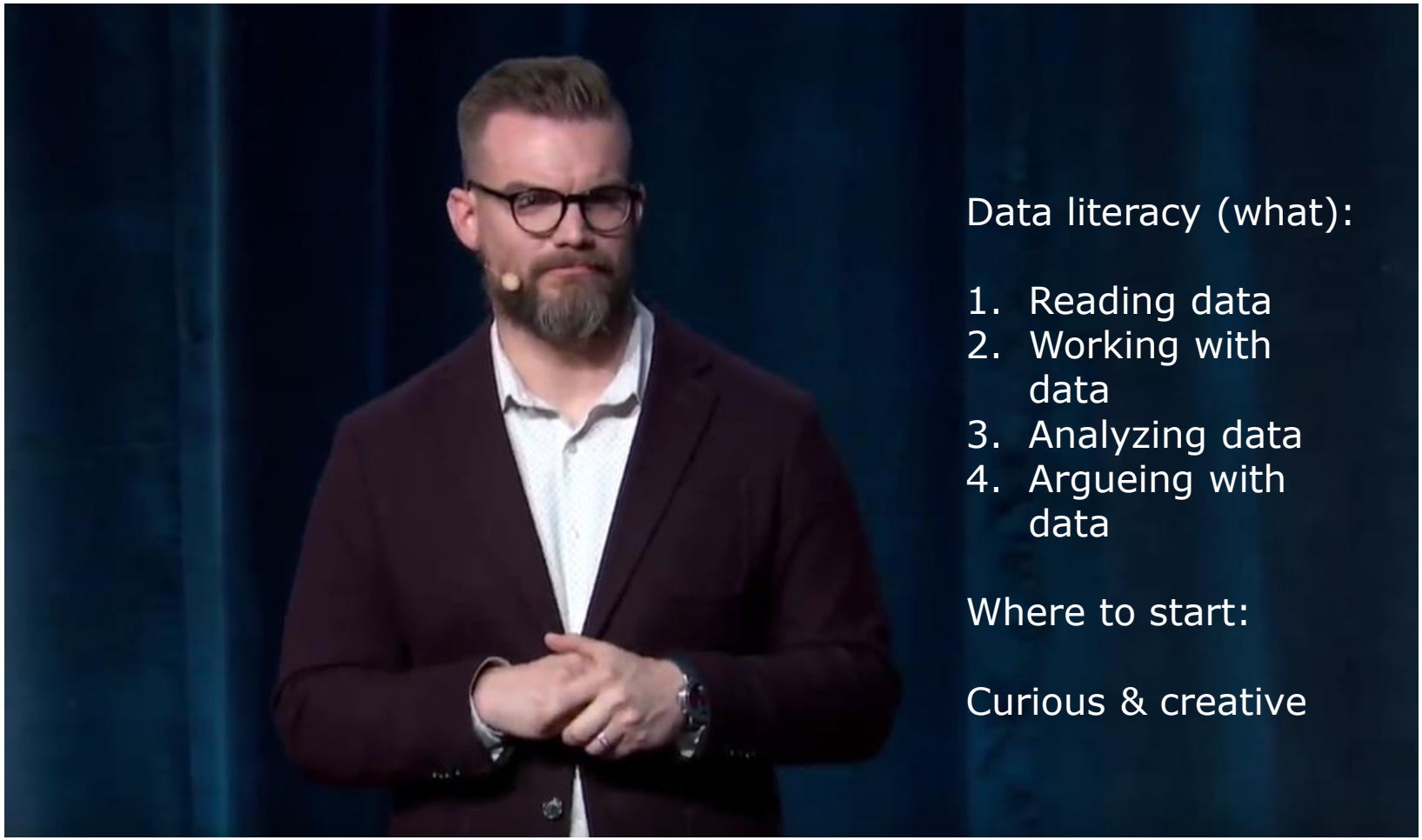
Data is a set of values of subjects with respect to qualitative or quantitative variables.

Data and information or knowledge are often used interchangeably; however data becomes information when it is viewed in context or in post-analysis.[1] While the concept of data is commonly associated with scientific research, data is collected by a huge range of organizations and institutions, including businesses (e.g., sales data, revenue, profits, stock price), governments (e.g., crime rates, unemployment rates, literacy rates) and non-governmental organizations (e.g., censuses of the number of homeless people by non-profit organizations).

Data is measured, collected and reported, and analyzed, whereupon it can be visualized using graphs, images or other analysis tools. Data as a general concept refers to the fact that some existing information or knowledge is *represented* or *coded* in some form suitable for better usage or processing. *Raw data* ("unprocessed data") is a collection of numbers or characters before it has been "cleaned" and corrected by researchers. Raw data needs to be corrected to remove outliers or obvious instrument or data entry errors (e.g., a thermometer reading from an outdoor Arctic location recording a tropical temperature). Data processing commonly occurs by stages, and the "processed data" from one stage may be considered the "raw data" of the next stage. Field data is raw data that is collected in an uncontrolled "in situ" environment. Experimental data is data that is generated within the context of a scientific investigation by observation and recording. Data has been described as the new oil of the digital economy.[2][3]

https://en.wikipedia.org/wiki/Data

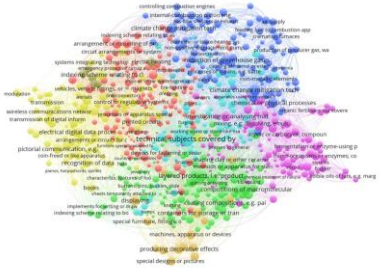
Lektie: https://youtu.be/8ovyQZ_Z8Xs



Den praktiske virkelighed fra HELP DESK´en

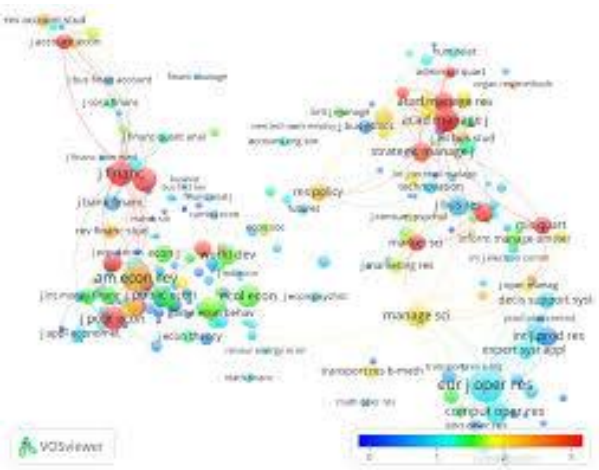
ShareLaTeX

API



Case 1 - søgning fra Scopus til VosViewer

	Authors	Paper Title	Year	Source title
1.	Markmann C. et al.	A Delphi-based risk analysis - Identifying and assessing future challenges for supply chain security in a multi-stakeholder environment	2013	Technological Forecasting and Social Change
2.	Schippl J.	Assessing the desirability and feasibility of scenarios on eco-efficient transport: a heuristic for efficient stakeholder involvement during foresight processes	2016	Foresight
3.	Gravagnuolo A., et al.	Assessment of waterfront attractiveness in port cities - Facebook 4 Urban Facelifts	2015	International Journal of Global Environmental Issues
4.	Karger C.R.	Citizen scenarios for the future of personalized medicine: A participatory scenario process in Germany	2013	International Journal of Interdisciplinary Social and Community



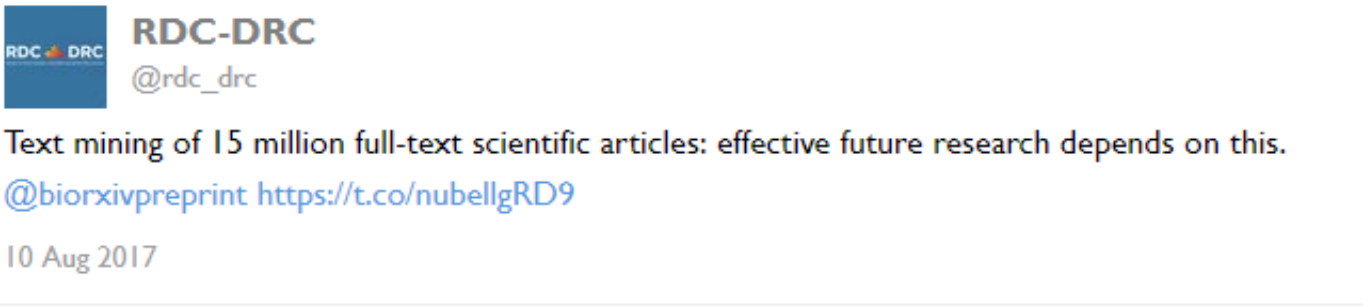
30 artikler via tabel i Outlook

Scopus → CSV

Case 2 - Text og Data Mining adgang

Fra licens til formidling

Text mining of 15 million full-text scientific articles
David Westergaard, Hans-Henrik Stærfeldt, Christian Tønsberg, Lars Juhl Jensen, Søren Brunak
doi: https://doi.org/10.1101/162099
Now published in *PLOS Computational Biology*
doi: 10.1371/journal.pcbi.1005962



+ input fra andre eksperter om metode

Egon Willighagen • 2 years ago

Review of the abstract:

Text mining 15 million articles is mostly just putting a nice cluster in place and doing some useful coordination. The results are more interesting, and I would suggest to have the title reflect that. For example, "15 million articles recovered interaction data for 60M protein, gene pairs"? The abstract itself does not really give a lot of detail about the method used, nor about the key findings of differences in old versus new literature. Did you find the literature to reflect the rise and decline of experimental protocols to measure these things? Also, would love to read in the abstract what was done with the resulting data. Was it release as Open Data, and/or via a SPARQL end point? Did it get integrated in existing databases, e.g. Ensembl or UniProt?

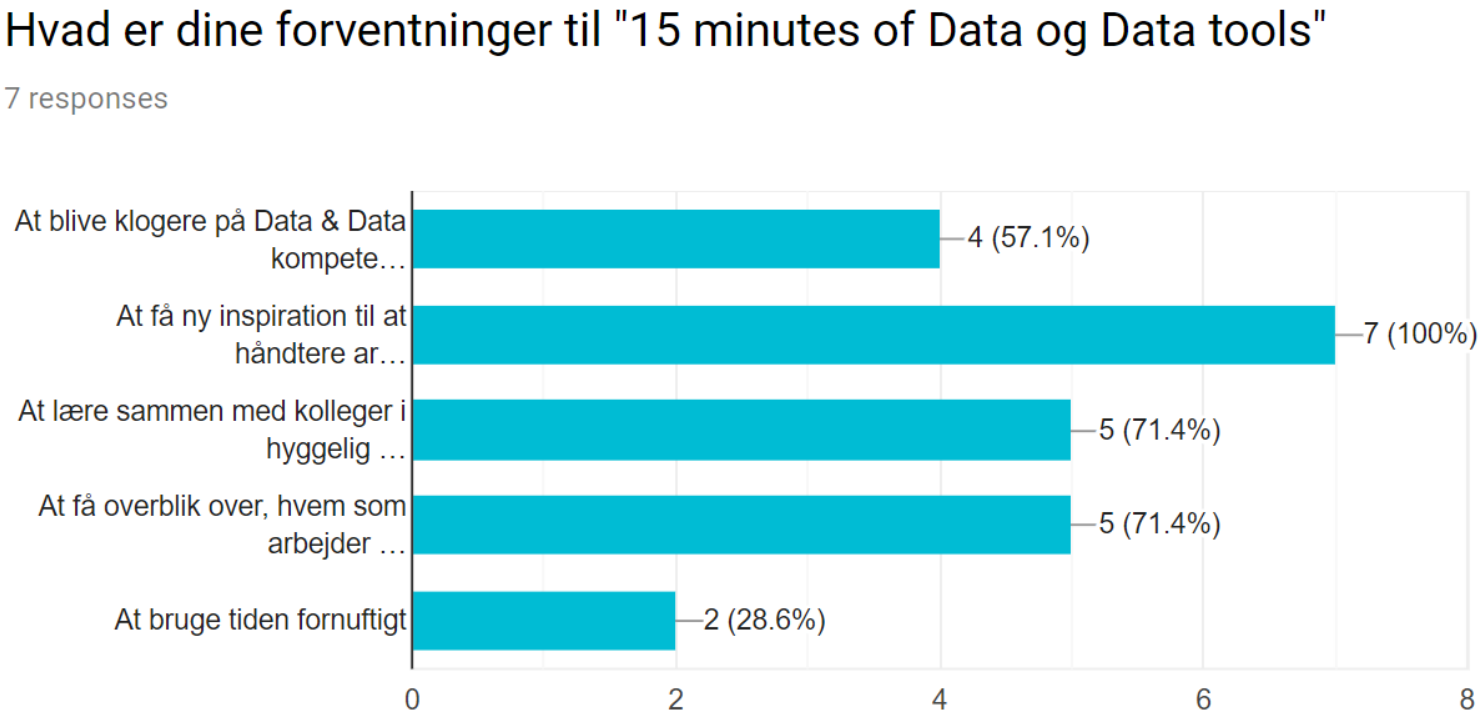
That said, looking forward to reading the full paper during my holidays!

Argumenter for at lære det☺

Liberating Data: How libraries and librarians can help researchers with text and data mining.

https://blogs.lse.ac.uk/impactofsocialsciences/2016/07/12/how-libraries-and-librarians-can-help-with-text-and-data-mining/

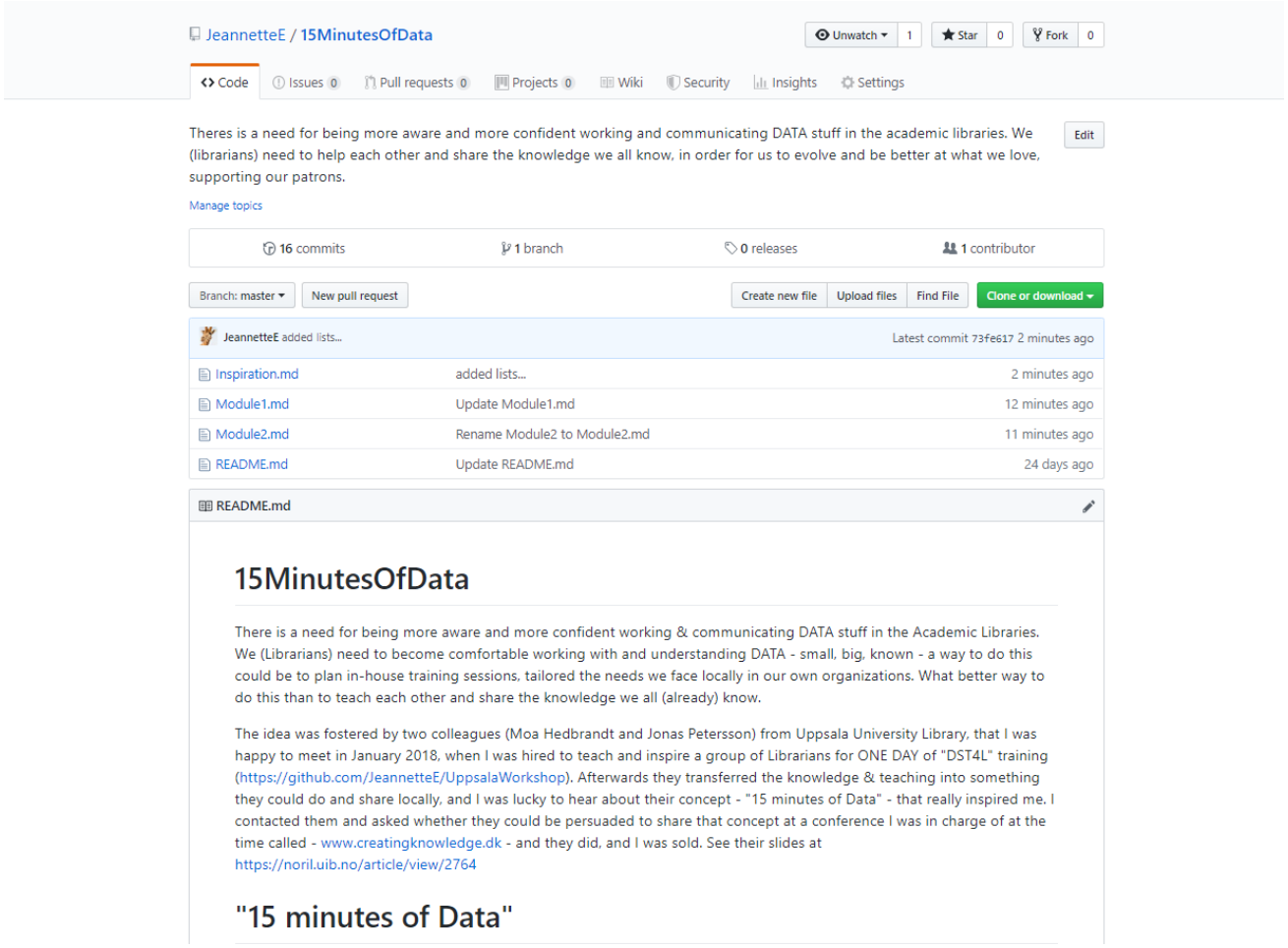
Vi skal derfor i gang – og



Udkast til planen for efteråret 2019

Indhold	Forberedelse	Inspiration/tools
Reading Data – *kende data, kende egne systemer og muligheder*	Videoer og/eller artikler	Links til værktøjer
Working with data – *rense data, finde "outliers" i mindre data set*	Videoer og/eller artikler	Links til værktøjer
Analyzing data – *visualisere, sammenholde, samle, dele*	Videoer og/eller artikler	Links til værktøjer
Arguing with data – *finde de gode historier, videndele erfaringer*	Videoer og/eller artikler	Links til værktøjer

Deles via mail☺ og GitHUB

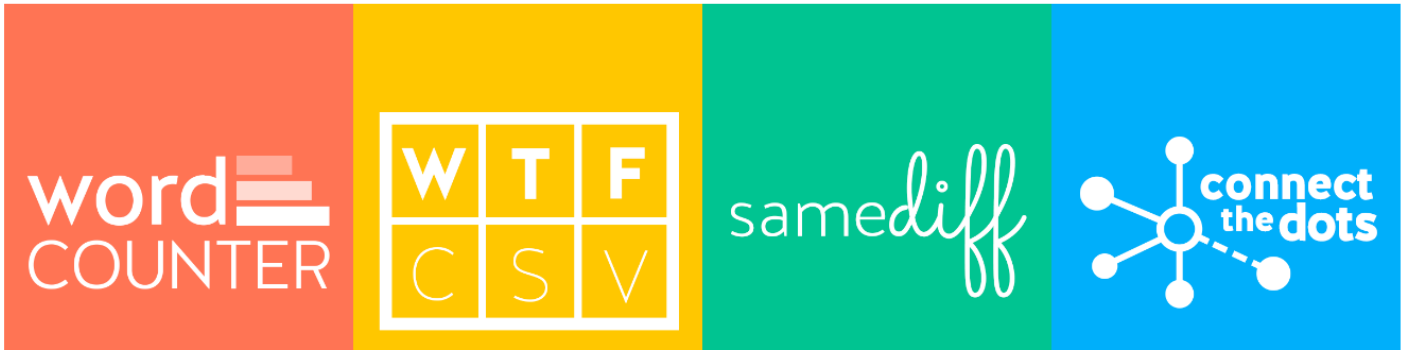


https://github.com/JeannetteE/15MinutesOfData

Dagens Hands ON

data BASIC.io

DataBasic is a suite of easy-to-use web tools for beginners that introduce concepts of working with data. These simple tools make it easy to work with data in fun ways, so you can learn how to find great stories to tell.



https://databasic.io/en/

Input til JEEK til næste gang d. 5. september