

15-30 minutes of

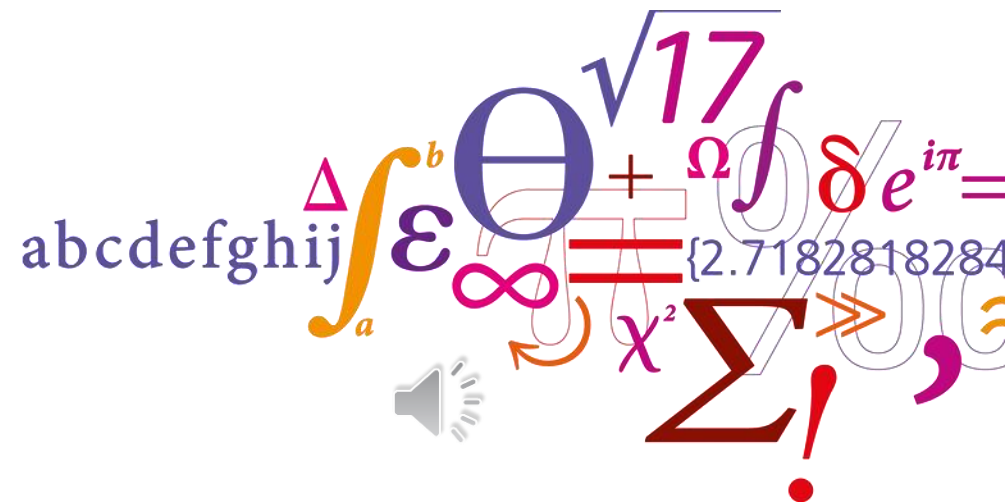
Data & Data Science Tools

Videndeling, inspiration, hænder-på

Koordineret og eksekveret af
DTU Bibliotek

V Jeannette Ekstrøm
Senior Information Officer
jeek@dtu.dk

Twitter: @JEkstroem



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"



© 2017 DIRK L. TIEDE - WWW.DIRKTIEDE.COM

THIS WORK IS LICENSED UNDER A CREATIVE COMMONS
ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE 3.0 UNPORTED LICENSE



Ideen inspireret fra CreatingKnowledge 2018



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.....

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



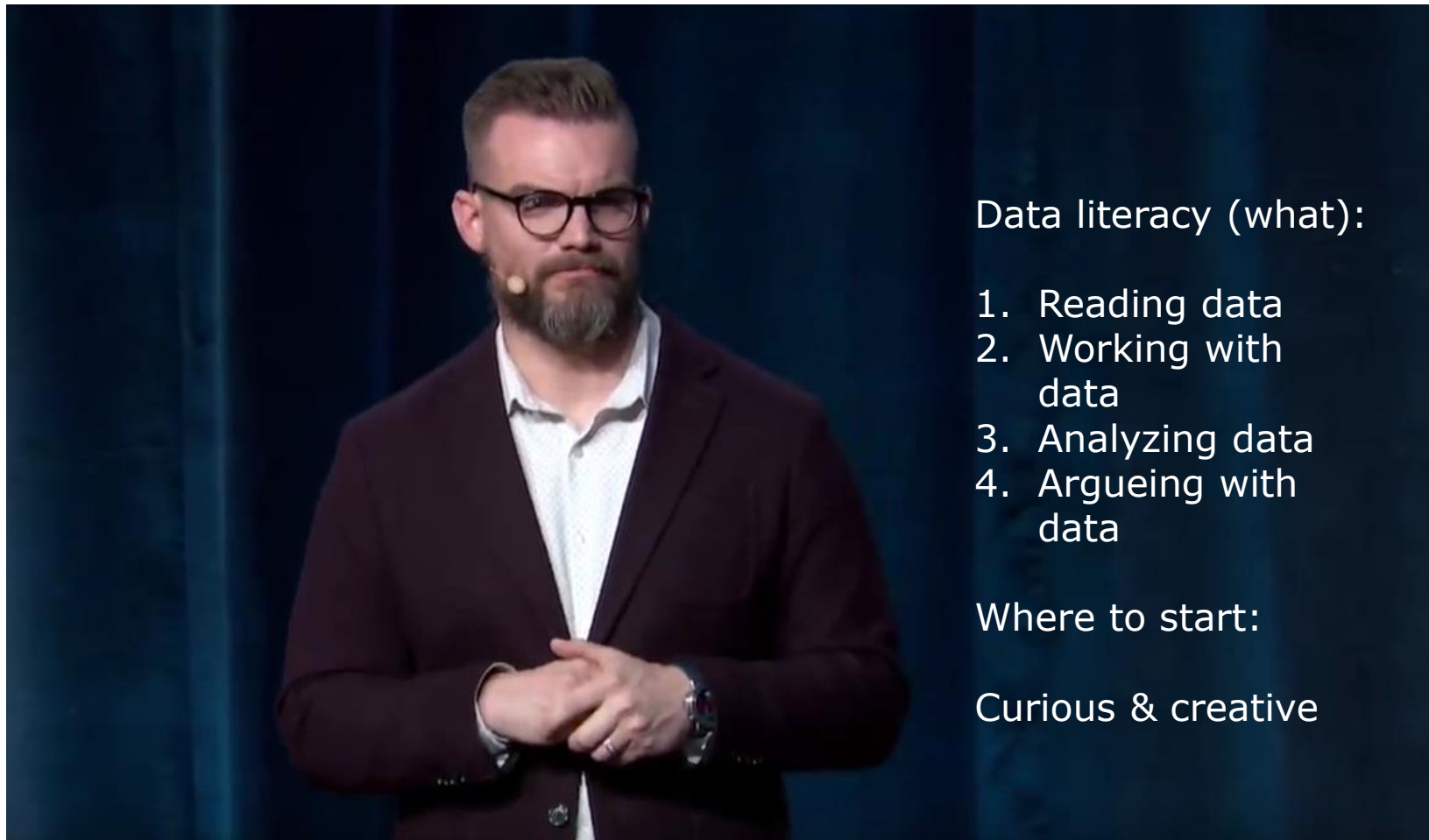
Data is a set of values of subjects with respect to [qualitative](#) or [quantitative variables](#).

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



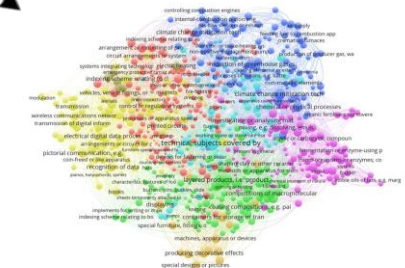
Data literacy (what):

1. Reading data
2. Working with data
3. Analyzing data
4. Argueing with data

Where to start:

Curious & creative

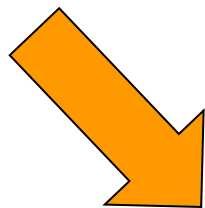
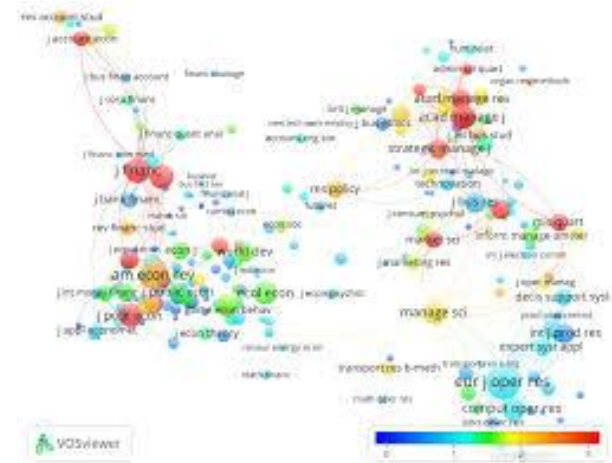
Den praktiske virkelighed fra HELP DESK'en



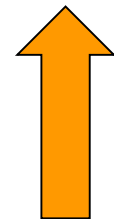
Case 1 - søgning fra Scopus til VosViewer

	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.	Schippel 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



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](https://doi.org/10.1371/journal.pcbi.1005962)



RDC-DRC

@rdc_drc

Text mining of 15 million full-text scientific articles: effective future research depends on this.

@biorxivpreprint <https://t.co/nubellgRD9>

10 Aug 2017

+ 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 ☺

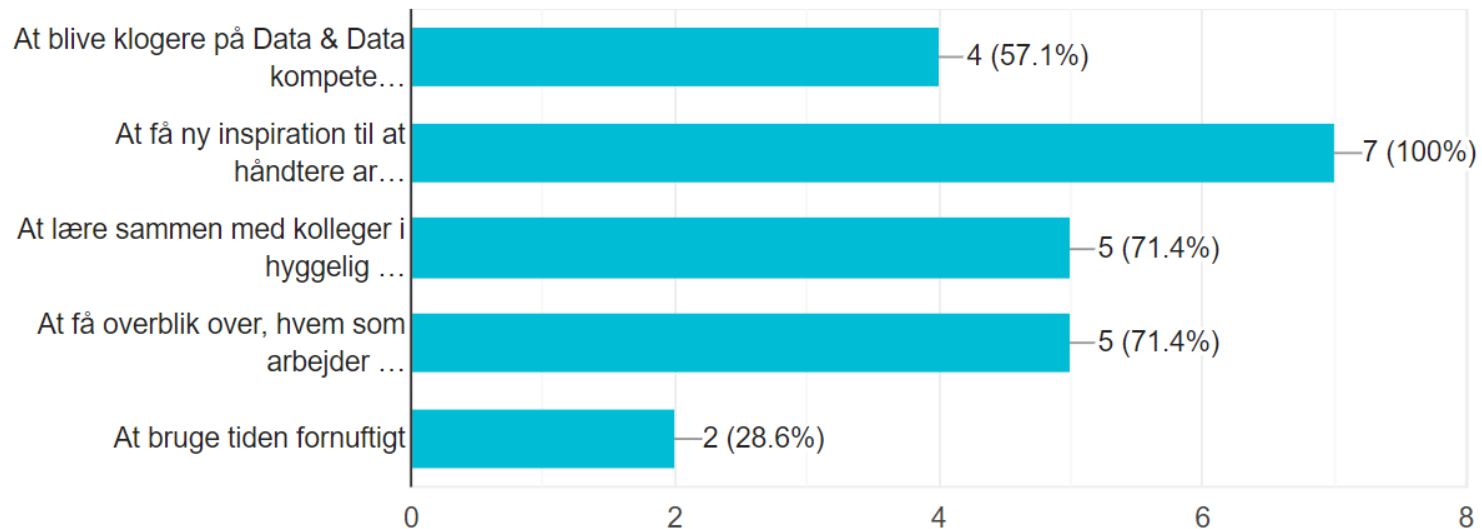
[Liberating Data: How libraries and librarians can help researchers with text and data mining.](https://blogs.lse.ac.uk/impactofsocialsciences/2016/07/12/how-libraries-and-librarians-can-help-with-text-and-data-mining/)

<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

Hvad er dine forventninger til "15 minutes of Data og Data tools"

7 responses



Udkast til planen for efteråret 2019

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

Deles via mail☺ og GitHub

JeannetteE / 15MinutesOfData
Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

There is a need for being more aware and more confident working and communicating DATA stuff in the academic libraries. We (librarians) need to help each other and share the knowledge we all know, in order for us to evolve and be better at what we love, supporting our patrons.

Manage topics

16 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find File Clone or download

JeannetteE added lists...	Latest commit 73fe617 2 minutes ago
Inspiration.md	added lists... 2 minutes ago
Module1.md	Update Module1.md 12 minutes ago
Module2.md	Rename Module2 to Module2.md 11 minutes ago
README.md	Update README.md 24 days ago

README.md

15MinutesOfData

There is a need for being more aware and more confident working & communicating DATA stuff in the Academic Libraries. We (Librarians) need to become comfortable working with and understanding DATA - small, big, known - a way to do this could be to plan in-house training sessions, tailored the needs we face locally in our own organizations. What better way to do this than to teach each other and share the knowledge we all (already) know.

The idea was fostered by two colleagues (Moa Hedbrandt and Jonas Petersson) from Uppsala University Library, that I was happy to meet in January 2018, when I was hired to teach and inspire a group of Librarians for ONE DAY of "DST4L" training (<https://github.com/JeannetteE/UppsalaWorkshop>). Afterwards they transferred the knowledge & teaching into something they could do and share locally, and I was lucky to hear about their concept - "15 minutes of Data" - that really inspired me. I contacted them and asked whether they could be persuaded to share that concept at a conference I was in charge of at the time called - www.creatingknowledge.dk - and they did, and I was sold. See their slides at <https://noriluib.no/article/view/2764>

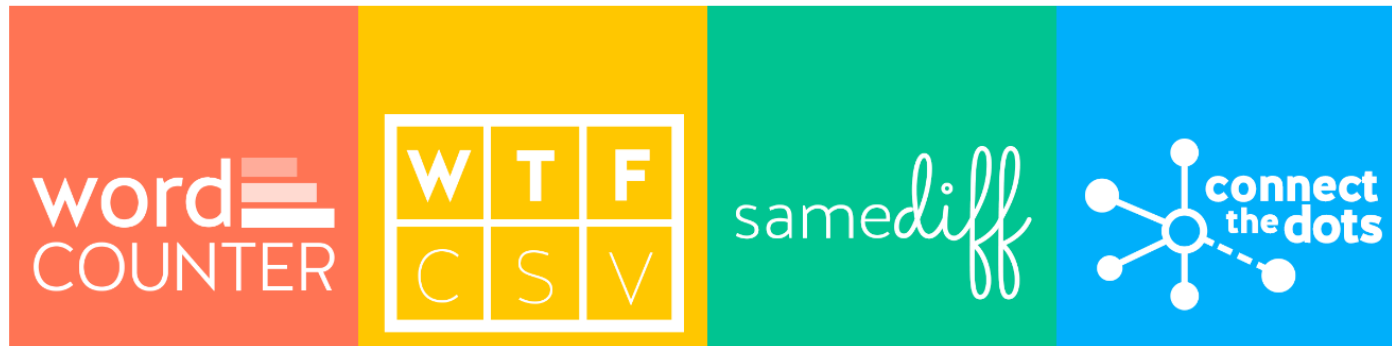
"15 minutes of Data"

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

Dagens Hands ON



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



Modtages gerne