Digital media lab

Digital data in academia

DiMeLab
Digital Media Lab · Roskilde University

Why the Digital Media Lab?

- Research current technology and data practice
- Developing new methods from new data sources and infrastructure
- Understand the limitations and possibilities of digital data for academic and commercial purposes
- Enhancing critical reflection through technical insight

DiMeLab
Digital Media Lab · Roskilde University

Key services

- Homepage: https://digitalmedialab.ruc.dk/
- Tools
  - Twitter TCAT is running
  - Facebook tools via DMI
  - Instagram in test
  - YouTube via DMI
- Office hours every other Wednesday from 13-15 (Sander)
- Special events such as workshops and research seminars
- Supporting student projects at IKH

DiMeLab
Digital Media Lab · Roskilde University

Other digital labs (INT/DK)

- DMI, Amsterdam
  - wiki.digitalmethods.net/
- TANT, Aalborg (Kbh)
  - www.tantlab.aau.dk/
- ITU, København
  - ethos.itu.dk/
- AAU, Aarhus
  - www.digitalfootprints.dk/
- RUC, Roskilde
  - https://digitalmedialab.ruc.dk/


DATAFICATION

DiMeLab
Digital Media Lab · Roskilde University

Media revolution

DiMeLab
Digital Media Lab · Roskilde University

Data revolution

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

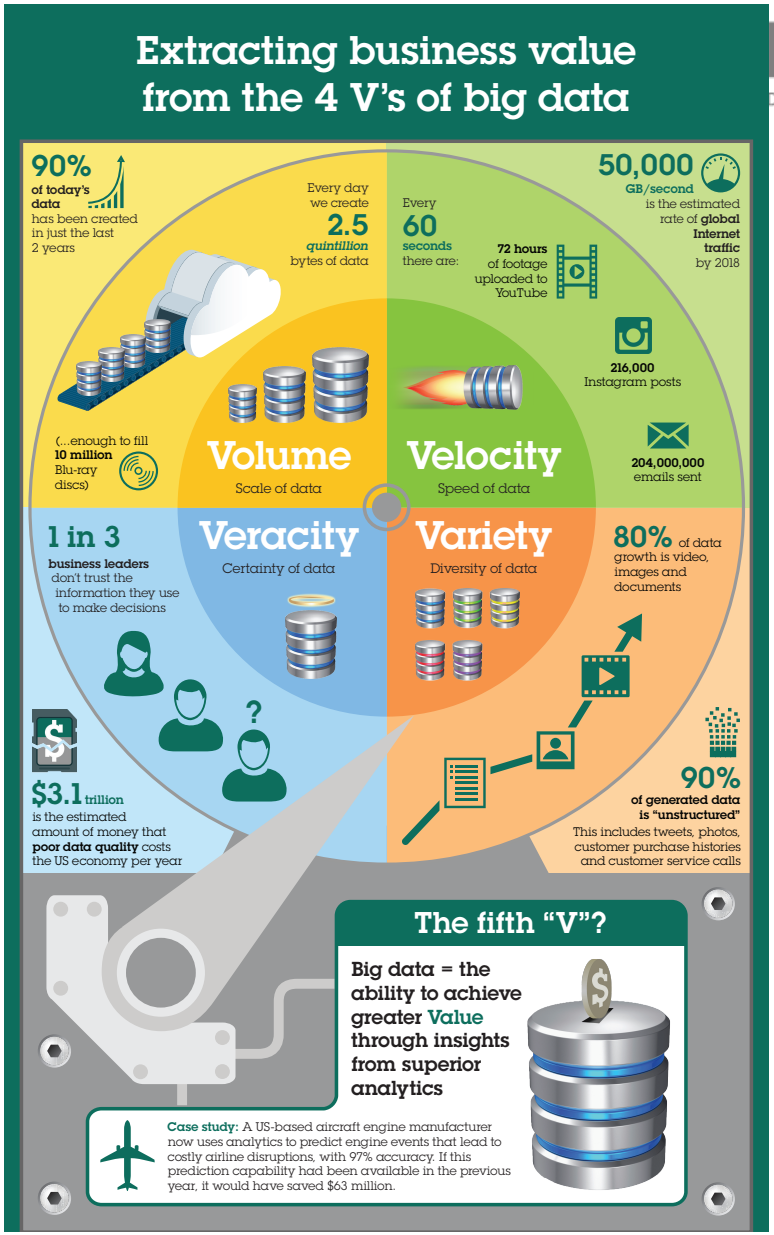
The data economy demands a new approach to antitrust rules

https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource

- "Data is just like crude. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc., to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value." (Palmer, 2006)
- http://ana.blogs.com/maestros/2006/11/data_is_the_new.html

Big data and 4V

- **V**olume
- **V**elocity
- **V**ariety
- **V**eracity
- (**V**alue?)



http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data

DiMeLab
Digital Media Lab · Roskilde University

Datafication

- To **datafy** a phenomenon is to put it in a quantified format so it can be tabulated and analyzed
- **Digitalization** is the process of converting analog information into the zeroes and ones of binary code so computers can handle it

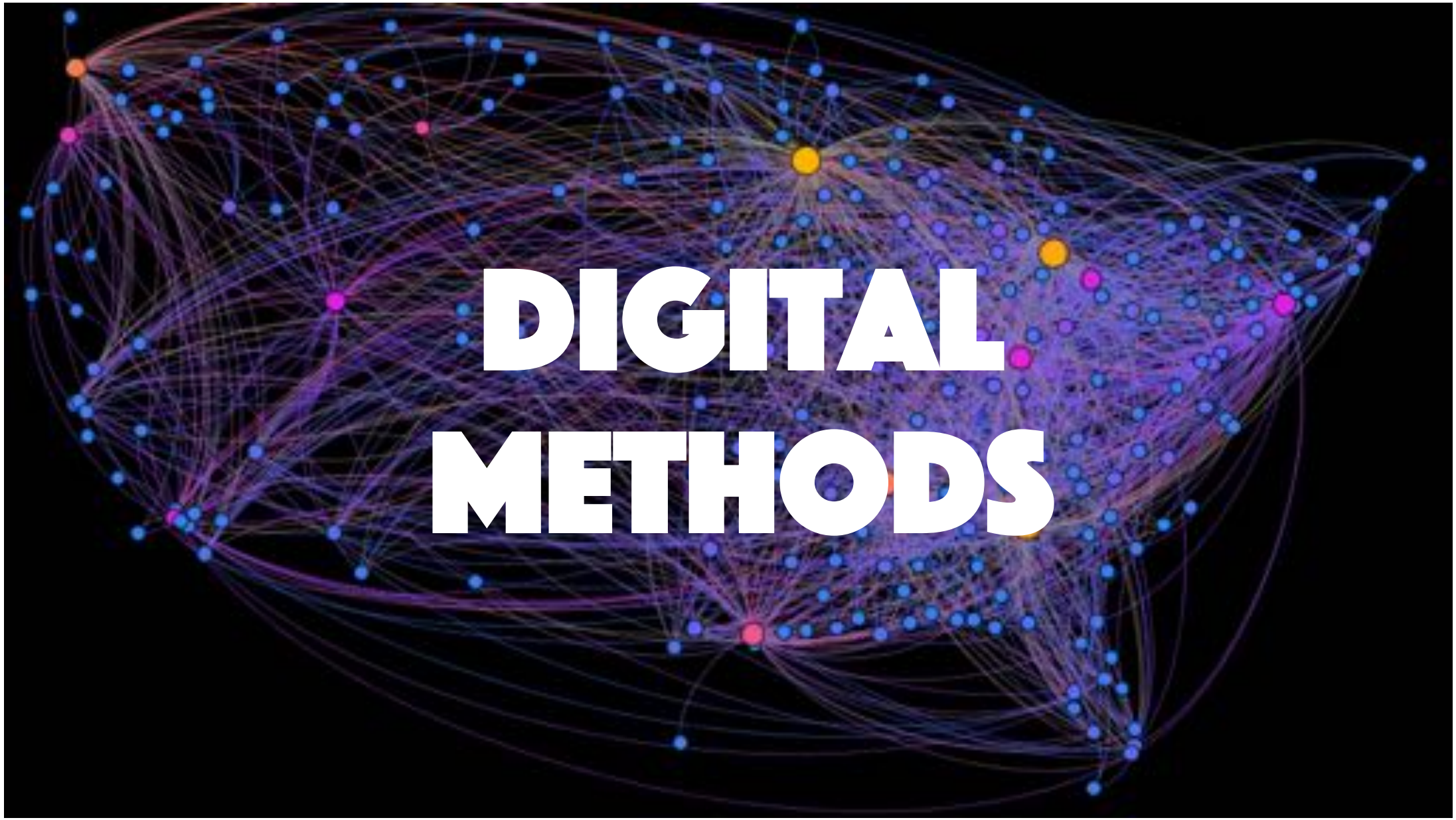
Mayer-Schonberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work and Think. London: John Murray.

DiMeLab
Digital Media Lab · Roskilde University

2018

GDPR
The General Data Protection Regulation

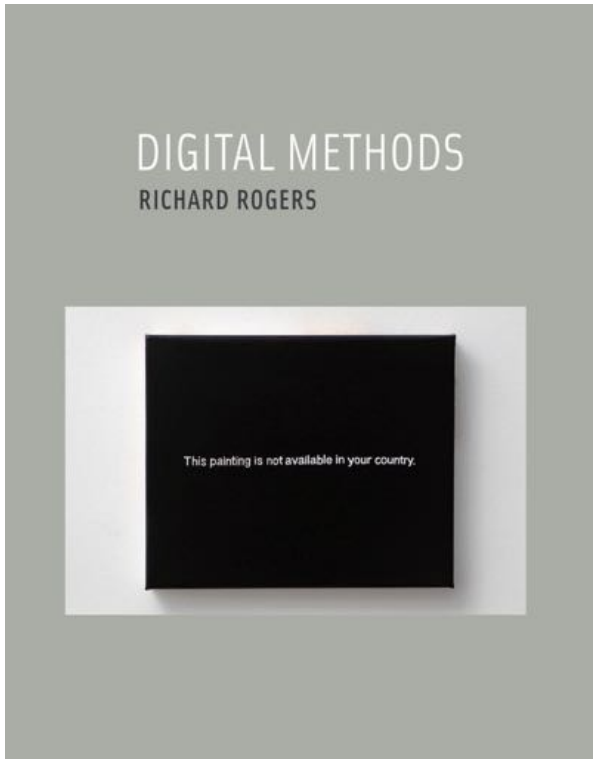
Cambridge Analytica


DIGITAL
METHODS

DiMeLab
Digital Media Lab · Roskilde University

Digital methods

”Digital methods is a term that seeks to capture a recent development in Internet-related research, summarized as approaches to the web as data set. Joining a larger computational turn in the social sciences and the digital humanities, it asks a series of questions about the quality of web data, the productivity of online collection and analytical methods, and ultimately the prospects of having the web serve as a site for grounding findings. When may the web become the baseline for findings about social change?” (Rogers, 2015)



Rogers, R. (2015). Digital Methods for Web Research. In Robert A. Scott & S. M. Kosslyn (Eds.), Emerging Trends in the Behavioral and Social Sciences. Hoboken, NJ: Wiley. doi:10.1002/9781118900772

Rogers, R. (2013). Digital methods. Cambridge, Massachusetts: The MIT Press.

DiMeLab
Digital Media Lab · Roskilde University

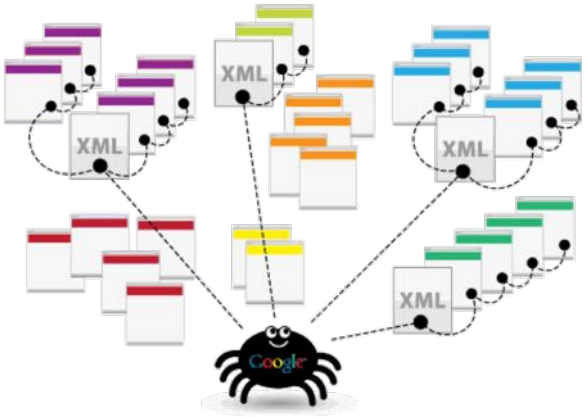
Digital Methods Initiative

- ”The Digital Methods Initiative is a contribution to doing research into the "natively digital". [...], the focus is on how methods may change, however slightly or wholesale, owing to the technical specificities of new media.”
  - https://wiki.digitalmethods.net/Dmi/WebHome

DiMeLab
Digital Media Lab · Roskilde University

Digital collection or analysis

- Digital collection
  1. Scraping
  2. Crawling (Google spider)

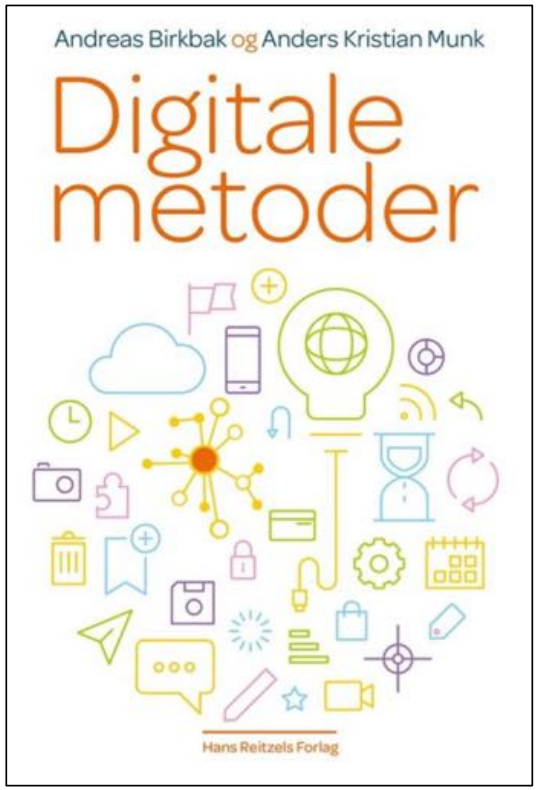


- Digital analysis
  - Traditional methods
    - Observation
    - Content analysis
    - Discourse analysis
    - …
  - Digital methods
    - Automatic
    - Software defined
    - Quantitative or mixed-methods

http://medieogkommunikationsleksikon.dk/digitale-metoder/

DiMeLab
Digital Media Lab · Roskilde University

Digital methods in Danish (2017)




DATA
COLLECTION

DiMeLab
Digital Media Lab · Roskilde University

API according to text

- ”An API is an interface provided by an application that lets users interact with or respond to data or service requests from another program, other applications, or Web sites. APIs facilitate data exchange between applications, allow the creation of new applications, and form the foundation for the “Web as a platform” concept.” (Murugesan, 2007 in Helmond, 2015)

Helmond, A. (2015). The Platformization of the Web: Making Web Data Platform Ready. Social Media+ Society, 1(2), 2056305115603080.

DiMeLab
Digital Media Lab · Roskilde University

API according to Wikipedia

Technical

- **A**pplication **P**rogramming **I**nterface
- ”In computer programming, an **a**pplication **p**rogramming **i**nterface (API) is a set of subroutine definitions, protocols, and tools for building application software. In general terms, it is a set of clearly defined methods of communication between various software components.” (Wikipedia)
- https://en.wikipedia.org/wiki/Application_programming_interface

Less technical

- APIs are protocols that define how different types of software can communicate with each other
- APIs allow requests and give structured response
- Metaphorically APIs are like waiters that walks between the customer (client) and the kitchen (server)
- https://youtu.be/s7wmiS2mSXY

DiMeLab
Digital Media Lab · Roskilde University

APIs and academia

- ”Some social media companies make their data banks on users and usage patterns available through their APIs. Hence, the API is also an interface for researchers to collect data from a given social media service. Through small software scripts, researchers can access the API to retrieve, store, and manipulate digital traces left by the users of a service for further empirical analysis.” (Lomborg og Bechmann, 2014)

Lomborg, S., & Bechmann, A. (2014). Using APIs for Data Collection on Social Media. *The Information Society*, *30*(4), 256–265.

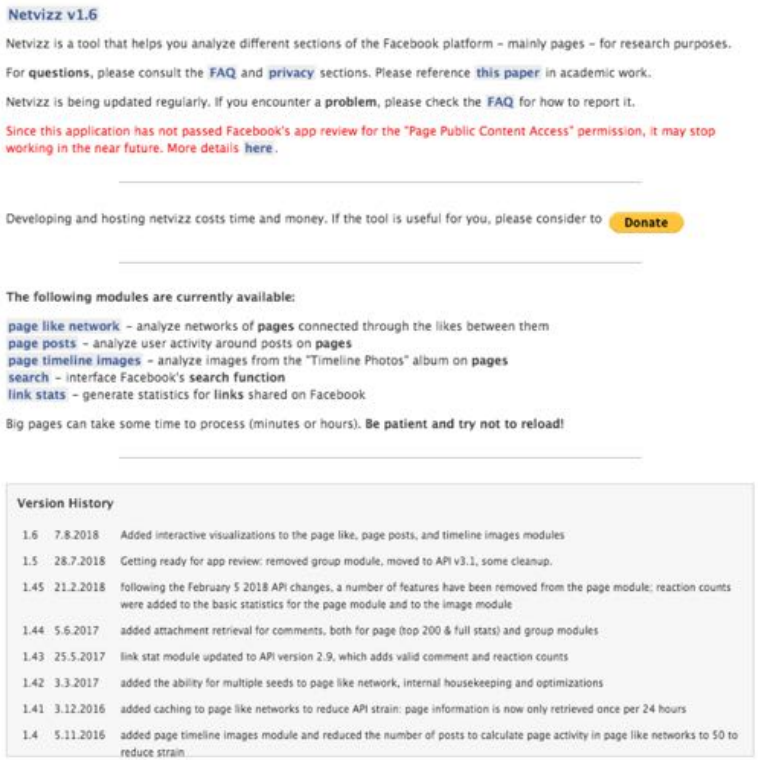
DiMeLab
Digital Media Lab · Roskilde University

Social media API interfaces/collection tools

	Text	Metrics	Comments	Visual	Network
Facebook (Netvizz)	X	X	X		
Twitter (TCAT)	X				X
Instagram (Instaload)	X			X	
YouTube (Data tool)	X	X	X		X

DiMeLab
Digital Media Lab · Roskilde University

Netvizz



Netvizz v1.6

Netvizz is a tool that helps you analyze different sections of the Facebook platform – mainly pages – for research purposes.

For **questions**, please consult the **FAQ** and **privacy** sections. Please reference **this paper** in academic work.

Netvizz is being updated regularly. If you encounter a **problem**, please check the **FAQ** for how to report it.

Since this application has not passed Facebook's app review for the "Page Public Content Access" permission, it may stop working in the near future. More details **here**.

Developing and hosting netvizz costs time and money. If the tool is useful for you, please consider to **Donate**

The following modules are currently available:

page like network – analyze networks of **pages** connected through the likes between them
page posts – analyze user activity around posts on **pages**
page timeline images – analyze images from the "Timeline Photos" album on **pages**
search – interface Facebook's **search function**
link stats – generate statistics for **links** shared on Facebook

Big pages can take some time to process (minutes or hours). **Be patient and try not to reload!**

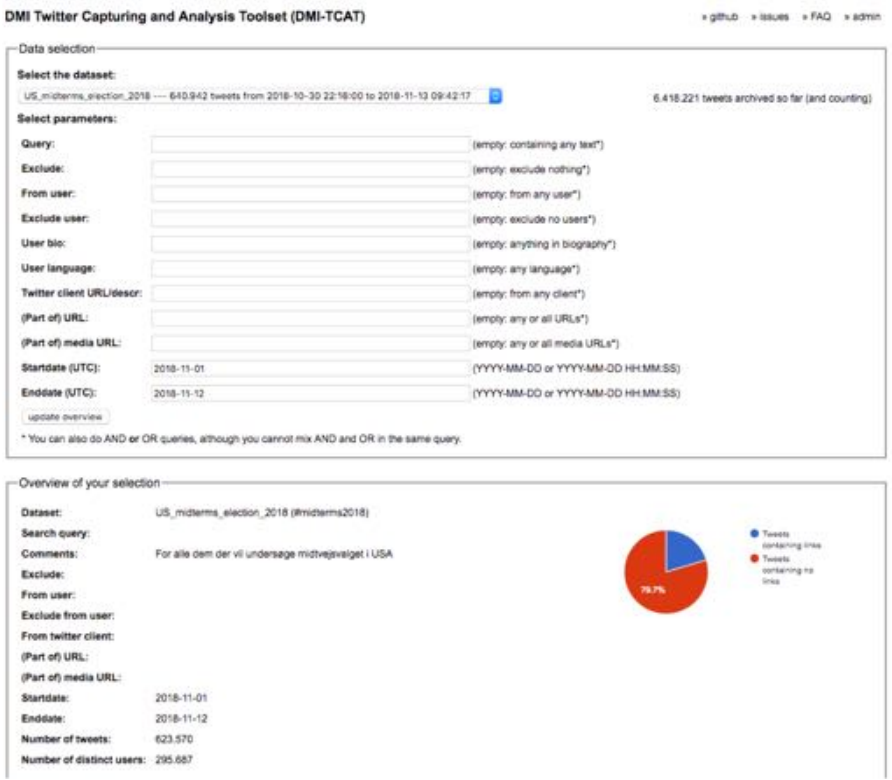
Version History

1.6	7.8.2018	Added interactive visualizations to the page like, page posts, and timeline images modules
1.5	28.7.2018	Getting ready for app review: removed group module, moved to API v3.1, some cleanup.
1.45	21.2.2018	following the February 5 2018 API changes, a number of features have been removed from the page module; reaction counts were added to the basic statistics for the page module and to the image module
1.44	5.6.2017	added attachment retrieval for comments, both for page (top 200 & full stats) and group modules
1.43	25.5.2017	link stat module updated to API version 2.9, which adds valid comment and reaction counts
1.42	3.3.2017	added the ability for multiple seeds to page like network, internal housekeeping and optimizations
1.41	3.12.2016	added caching to page like networks to reduce API strain: page information is now only retrieved once per 24 hours
1.4	5.11.2016	added page timeline images module and reduced the number of posts to calculate page activity in page like networks to 50 to reduce strain

- Pros
  - Popular platform
  - Engagement metrics
- Cons
  - Restrictive access (public pages)
  - Unreliable API

DiMeLab
Digital Media Lab · Roskilde University

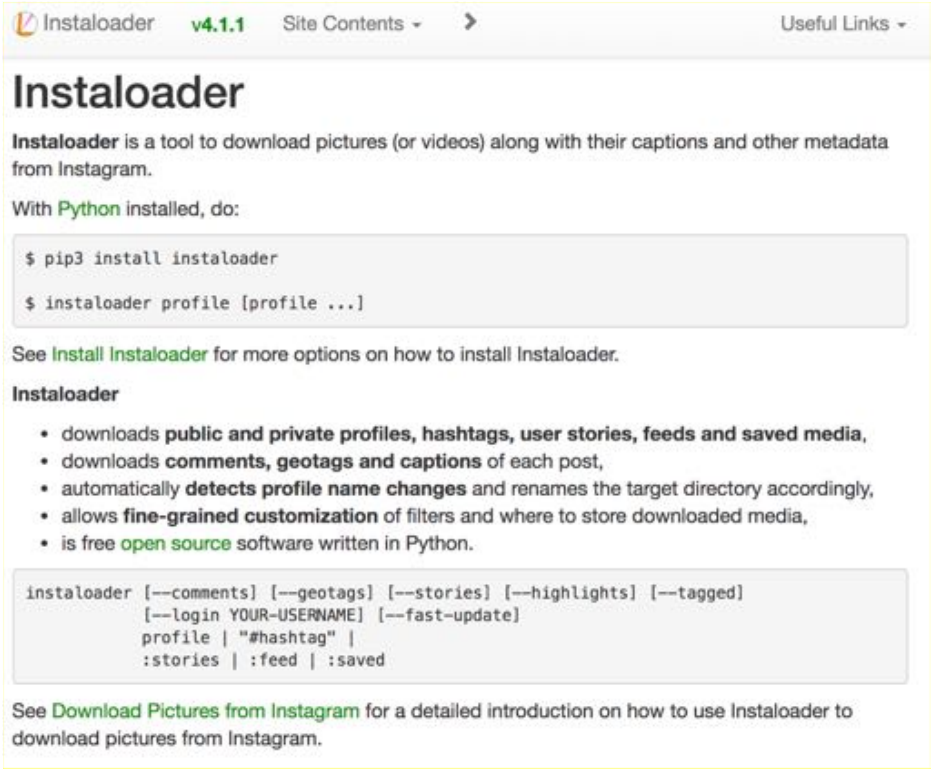
TCAT



- Pros
  - Reliable data access
  - Public by default platform
- Cons
  - Streaming API (not historical data)
  - 1% limitation sample
  - Metrics
  - Mostly media and politics

DiMeLab
Digital Media Lab · Roskilde University

Instagram


Instaloader v4.1.1 Site Contents Useful Links

Instaloader

Instaloader is a tool to download pictures (or videos) along with their captions and other metadata from Instagram.

With Python installed, do:

```
$ pip3 install instaloader

$ instaloader profile [profile ...]
```

See Install Instaloader for more options on how to install Instaloader.

Instaloader

- downloads **public and private profiles, hashtags, user stories, feeds and saved media**,
- downloads **comments, geotags and captions** of each post,
- automatically **detects profile name changes** and renames the target directory accordingly,
- allows **fine-grained customization** of filters and where to store downloaded media,
- is free open source software written in Python.

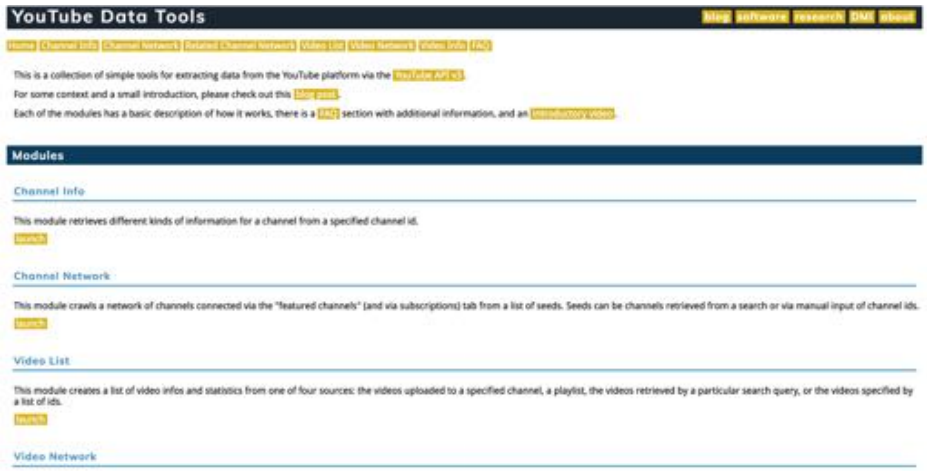
```
instaloader [--comments] [--geotags] [--stories] [--highlights] [--tagged]
            [--login YOUR-USERNAME] [--fast-update]
            profile | "#hashtag" |
            :stories | :feed | :saved
```

See Download Pictures from Instagram for a detailed introduction on how to use Instaloader to download pictures from Instagram.

- Pros
  - Popular platform with the youth
  - Public by default
- Cons
  - Not public API but scraping
  - Only collection from ind. profiles
  - No metrics
  - Messy output (can be fixed)

DiMeLab
Digital Media Lab · Roskilde University

YouTube



- Pros
  - API seems reliable (so far)
  - Popular platform
  - Networks
- Cons
  - Video centric

DiMeLab
Digital Media Lab · Roskilde University

D-I-Y

DiMeLab
Digital Media Lab · Roskilde University

What is the workshop about?

1. Collecting social media data via API
2. Getting to know your data: first steps
   - Getting from csv to Excel
   - Sorting data in Excel
   - Doing a time series analysis
3. Thinking about next steps: quant and qual

i. Not included in this workshop (but potential future workshops)
   - Inferential statistics and predictive analysis
   - NLP, AI and machine learning methods
   - Network analysis
   - Image recognition

- Please go to: digitalmedialab.ruc.dk/launch

DiMeLab
Digital Media Lab · Roskilde University

Output in CSV/TSV/TAB to Excel

- Delimiter-separated values
  - CSV: comma-separated values
  - TSV: Tab-separated values (.tab)
- Two-dimensional data divided into rows and columns in Excel
- Each row is a row but columns are defined through commas, tabs etc.

DiMeLab
Digital Media Lab · Roskilde University

Pivot tables in Excel