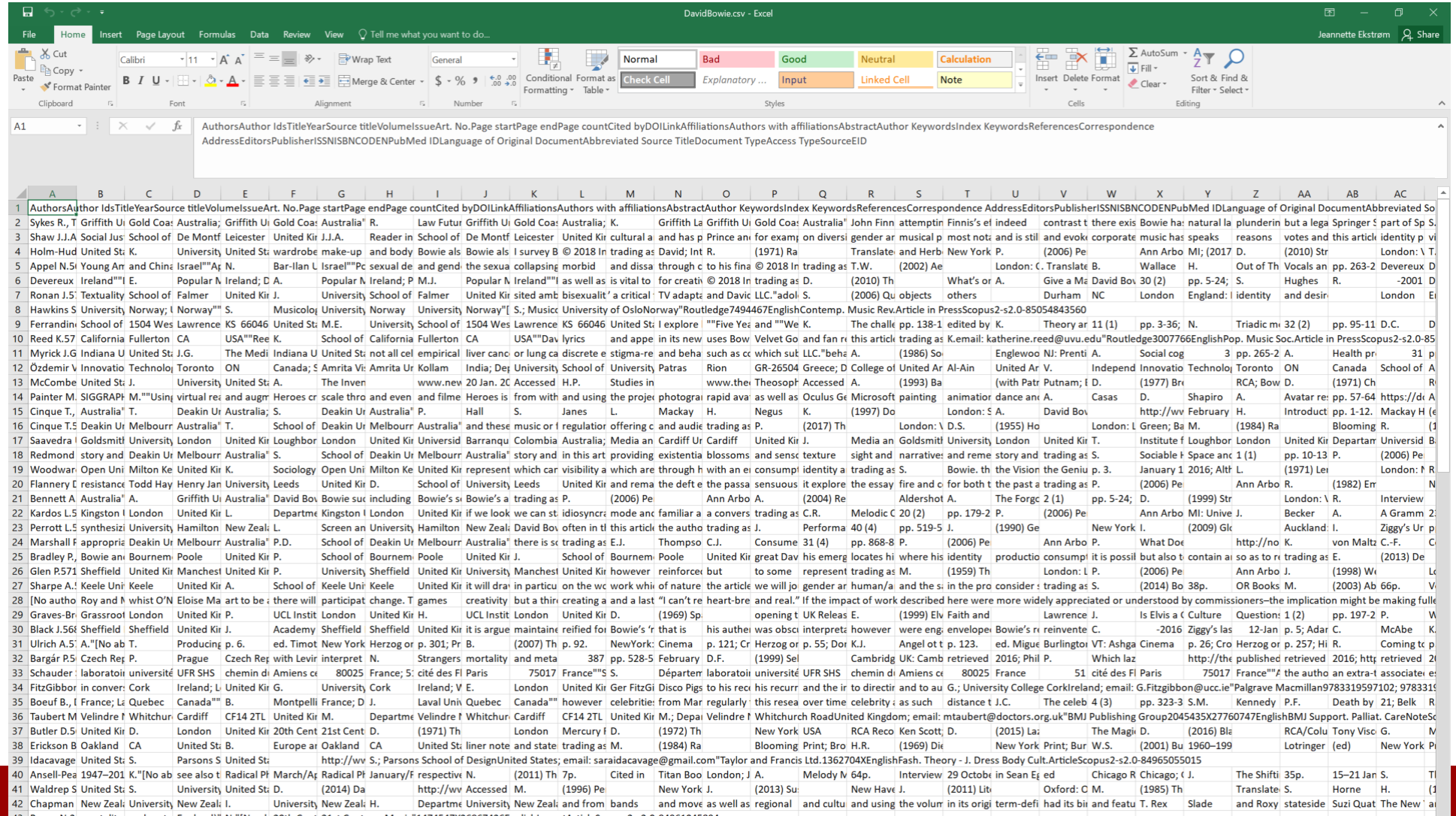
Session 3

15 minutes of Data

Opfølgning fra session 2…

http://scopus.com.proxy.findit.dtu.dk



.. I dag handler det om ”rensning og wrangling”

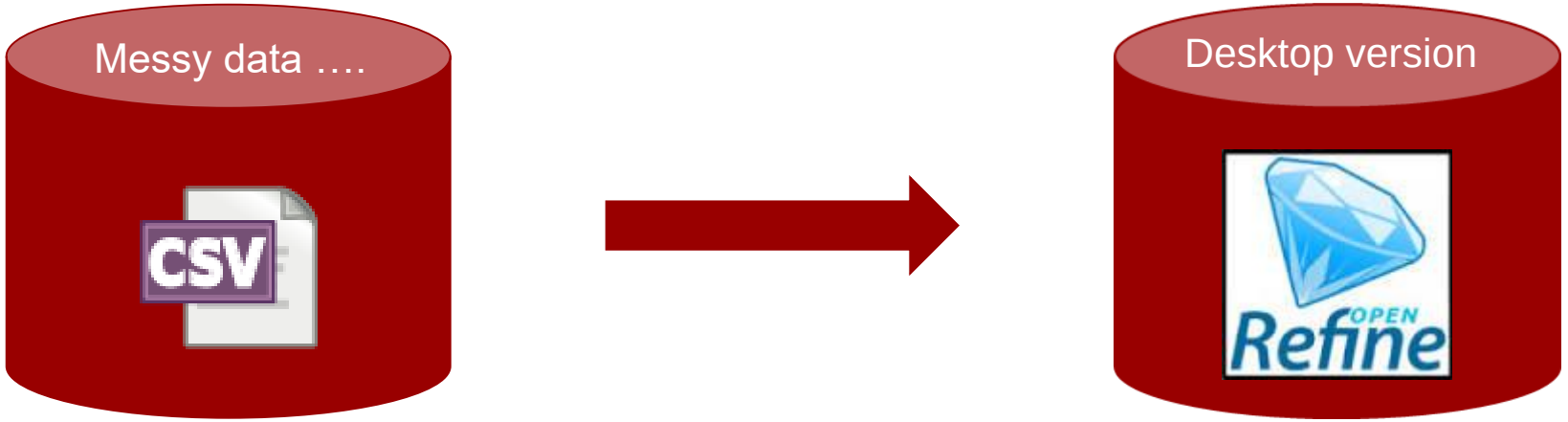
Adgang til ”dataset”	Rensning af dataset	Visualisering af dataset
A **data set** (or **dataset**) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question….	**Data cleansing** or **data cleaning** is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.	**Data visualization** is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewers of the images.
https://en.wikipedia.org/wiki/Data_set	https://en.wikipedia.org/wiki/Data_cleansing	https://en.wikipedia.org/wiki/Data_visualization
Scopus	OpenRefine	VOSviewer Visualizing scientific landscapes
http://scopus.com.proxy.findit.dtu.dk	http://openrefine.org/download.html	https://www.vosviewer.com/download

Hvad er “OpenRefine”

Whaat, really …both TSV, CSV, TXT, XML, RDF, JSON files – impressing, I like ☺

OpenRefine, formerly called **Google Refine** and before that **Freebase Gridworks**, is a standalone open source desktop application for data cleanup and transformation to other formats, the activity known as data wrangling.[3] It is similar to spreadsheet applications (and can work with spreadsheet file formats); however, it behaves more like a database.

Kilde: https://en.wikipedia.org/wiki/OpenRefine



Original fil ændres IKKE

Et stærkt værktøj, når det handler om “rensning”

- Eks fra https://librarycarpentry.org/lc-open-refine/01-introduction/index.html

- *Get an **overview of a data set***
- ***Resolve inconsistencies** in a data set, for example standardizing date formatting*
- *Help you **split data** up into more granular parts, for example splitting up cells with **multiple authors into separate cells***
- *Match local data up to other data sets, for example in matching local subjects against the Library of Congress Subject Headings*
- ***Enhance a data** set with data from other sources*

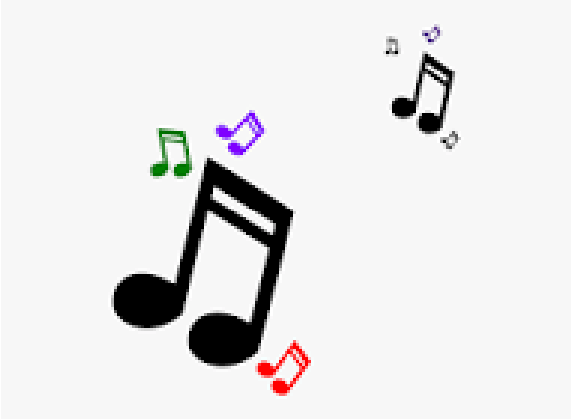
- Where you have a list of names or terms that differ from each other but refer to the same people, places or concepts. For example:

Data you have	Desired data
London	London
London]	London
London,]	London
london	London

credit: https://flic.kr/p/2aB3ABj

Todays assignment..

….. Simon/Jeannette says…



https://www.clipartwiki.com/iclip/imhw_colorful-music-note-clip-art-free-clipart-images/

- Open ”OpenRefine”
- Import your *.csv file from last session

- Click **NEXT**

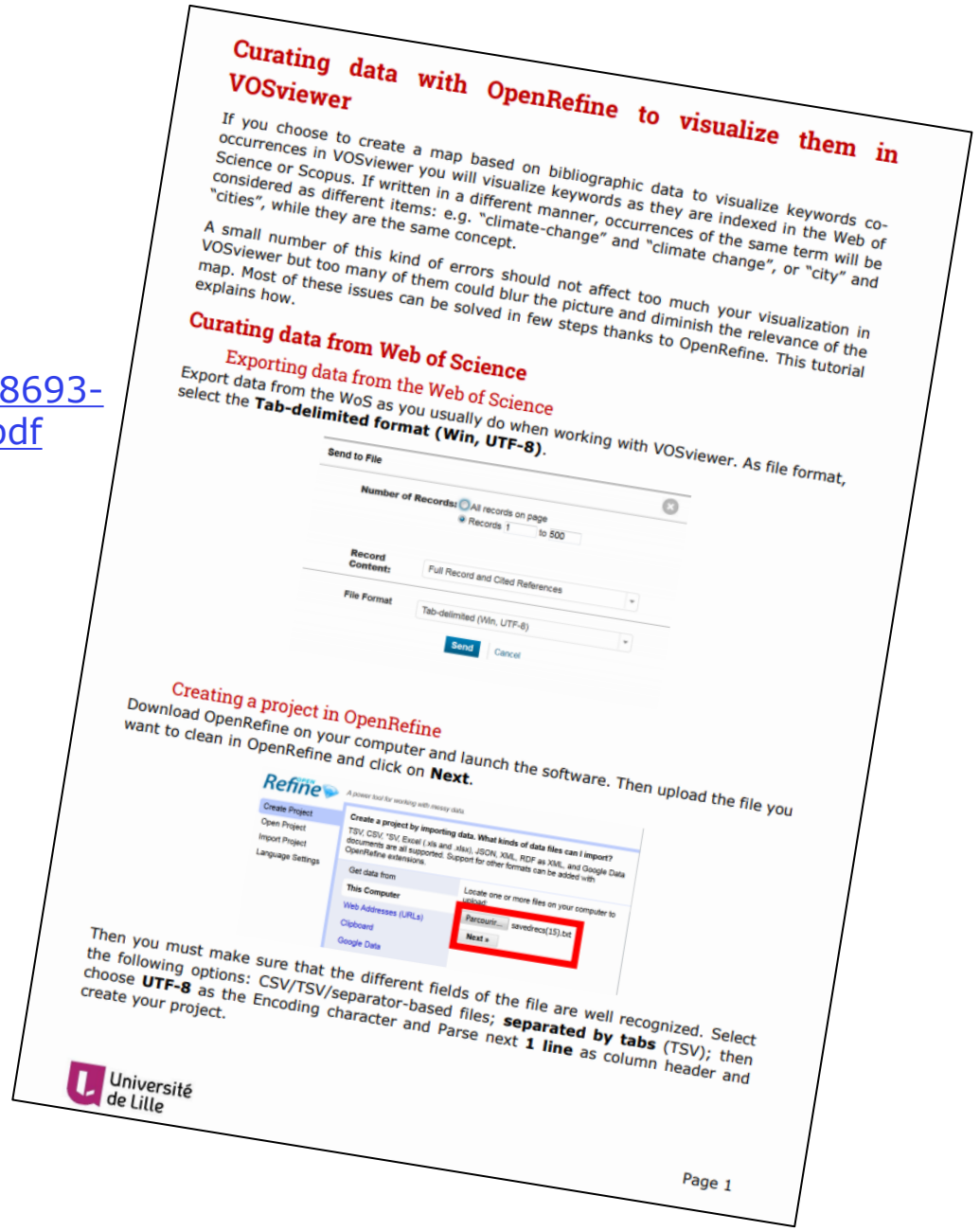
Select:

- CSV/TSV/separator-based files; choose CSV (comma)
- & UTF-8 (as character encoding)

& follow the guide ☺

God guide fra Universite de Lille

https://ged.univ-lille.fr/nuxeo/nxfile/default/9b2701b4-7b8f-4754-8693-7072ee219706/blobholder:0/tutorial_openrefine_vosviewer_feret.pdf



Curating data with OpenRefine to visualize them in VOSviewer

If you choose to create a map based on bibliographic data to visualize keywords co-occurrences in VOSviewer you will visualize keywords as they are indexed in the Web of Science or Scopus. If written in a different manner, occurrences of the same term will be considered as different items: e.g. "climate-change" and "climate change", or "city" and "cities", while they are the same concept.

A small number of this kind of errors should not affect too much your visualization in VOSviewer but too many of them could blur the picture and diminish the relevance of the map. Most of these issues can be solved in few steps thanks to OpenRefine. This tutorial explains how.

Curating data from Web of Science

Exporting data from the Web of Science

Export data from the WoS as you usually do when working with VOSviewer. As file format, select the **Tab-delimited format (Win, UTF-8)**.

Creating a project in OpenRefine

Download OpenRefine on your computer and launch the software. Then upload the file you want to clean in OpenRefine and click on **Next**.

Then you must make sure that the different fields of the file are well recognized. Select the following options: CSV/TSV/separator-based files; **separated by tabs** (TSV); then choose **UTF-8** as the Encoding character and Parse next **1 line** as column header and create your project.

Université de Lille

Page 1

Supplerende information, skulle I ønske mere…

- Online kursus fra Library Carpentry
  https://librarycarpentry.org/lc-open-refine/

- Eller en “gammel ven” Thomas Padilla
  http://thomaspadilla.org/dataprep/