# Assignment 2 Instructions

Margherita Fantoli

Introduction to Digital Humanities 2022-2023
`margherita.fantoli@kuleuven.be`

## 1 Introduction

While Assignment 1 had the goal to familiarize you with the process of data cleansing and enriching, Assignment 2 aims at letting you experiment with data visualization and analysis. The assignment is individual and will count for 40% of the final grade. **The deadline is January 9th, 12h, and the assignment should be submitted on Toledo**.

- You will be asked to deal with tools/topics that have only been briefly mentioned in class. This means that your technical results/in-depth previous knowledge won't be directly evaluated

- Your freedom is large, but don't get lost in it! The point is not to embark on challenging exercises – unless it is a priority for you – but to make targeted use of what you already know or would like to exercise upon.

**Bonus point**: You can learn to use LateX (a markup language often used in the hard sciences and programming community to deliver papers and abstracts) for your report. A nice blogpost of why humanities scholars should consider LateX is found here: `https://www.overleaf.com/blog/636-guest-blog-post-latex-for-the-humanities`. A simple introduction: `https://latex-tutorial.com/tutorials/`. but many others are available online. I especially recommend using Overleaf (`https://www.overleaf.com/` and `https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes`). You might consider using existing

templates, that can be found here: `https://www.latextemplates.com/`, `https://fr.overleaf.com/latex/templates/tagged/humanities`.This document was generated using this template: `https://www.overleaf.com/latex/templates/dh-intro-template/fyydynbrywqr`. **In case you proceed with LateX, please attach both the PDF and the tex file.**

# 2  Assignment description

Use your preferred favorite language/analysis tool to extract information for the dataset you used at *assignment 1*. Information can be: descriptive statistics, linguistic features, Social Network Analysis. Here are some examples: Number of different people/publishers, mapping and analysis of the distribution of the places of publication, counting of the frequencies of publications of a certain author/volume, linguistic analysis of the titles (most frequent terms, length, evolution across time/cities, language). You can apply network analysis to the people/entities involved in the publications. Try to exploit the linking with Wikidata to add more features to your analysis.

Tools that can be used are, but are not limited to: Python (esp. Pandas), R, Gephi for visualization of the graph (after having created with other tools, e.g. Python, the list of nodes and edges or the adjacency matrix), Tableau Public, Nodegoat, Excel (but showing profieciensy!) etc.

**Deliverables**

- Report (PDF, max 6 pages) with 1. Description of the tool(s) used 2. Results (and discussion) of the data analysis 3. Assessment on whether the structure/cleaning status of the dataset was suited to the task or whether additional manipulations were necessary

- Data visualizations (if any, and if not included entirely in the report): you are free to chose what is the best format (Images, a link to an online page, other kinds of formats)

- If available, the code/files with the manipulations

**Assessment**   The evaluation will address the correctness, precision, and motivation of the analysis, the efficacy of visualization, and the effort in

information extraction and elaboration. The quality of the code will not be assessed.

**Some suggestions** An example of analysis of a similar dataset can be found here: `https://github.com/martinaverna/Magister-Dixit-Metadata-Exploration` (created by a former student of the Master in the frame of their internship!). This Open Access book can provide a lot of interesting ideas and useful code: exploit it! `https://www.humanitiesdataanalysis.org/`. Here you can find some more indications for data analysis in Python: `https://carpentries-incubator.github.io/python-humanities-lesson/07-putting-it-all-together/`. Here a possible pipeline for geo-referencing and mapping `https://programminghistorian.org/en/lessons/mapping-with-python-leaflet`. Resources about Gephi: `https://gephi.org/users/`
Enjoy your work and in the meantime