Recap Diary (2020.03.10)

Old to do list (before the choice to use percentages)
https://docs.google.com/spreadsheets/d/1id7c0WnE-U-o0Wf9kbMBr-I1MEpU61MLyFkQ0jc5WI c/edit#gid=0

June 2019

Following the same idea of ERIH PLUS, we created the authority list of DH journals.
Based on the answers received in the Google form to send at DH 2019.
https://gspinaci.github.io/diary/06.07.2019.pdf

July 2019

Cluster each db starting from the results from the results i get when i was in Leiden
~4k cluster and\or bigger cluster ~30k pubs. Quality function must be high
(reason why Dimension has more cluster)

Create and plot VosViewer map files with DH and AH score per cluster
https://gspinaci.github.io/diary/07.26.2019.pdf

August 2019

Spatial and temporal analysis of DH journals (iteration 1)
https://gspinaci.github.io/diary/08.02.2019.pdf

Refining clusters (AH+SS main field is still too large):
First list of DH journals created (from 12 answers) and getting coverage of DBs
Use the first list to run evaluation: the results show clusters with very low percentage of DH publications. Trying to increase this percentage.
We did it: adding print issn and changing the level of mega journals into nothing. Considering Exclusively and Significantly as DH. Marginally is not taken into account.
https://gspinaci.github.io/diary/08.23.2019.pdf

Zooming into WoS, where we already have three levels of granularity. The result is not comparable to a known field of study. Due to low number of DH publications the resulting links between clusters are equal.
https://gspinaci.github.io/diary/08.30.2019.pdf

September 2019

Writing paper for AIUCD 2020

Second iteration: take top 5 journals for each cluster with at least 5% DH. Manually assess the resulting list of 236 journals (without duplicates)

Plot the VosViewer maps of the second iteration.

https://gspinaci.github.io/diary/09.27.2019.pdf

The results of the Leiden algorithm are not satisfying (Dimensions has too many publications, trying to have ~4k clusters result in a very low function level)

The main fields of study are not correct. Cherry picking borderline clusters under AH, we got physic publications.
Asked Ludo and Nees what to do for increasing the precision. They told me we should focus on analyzing only what we need ie AH (**we are not focusing anymore on retrieving main field of studies**)

October 2019

We must understand what is the spatial and temporal coverage for DH journals (iteration2) over DBs.
Created csv files and gantt (temporal) histogram (spatial) charts.

The idea is to retrieve DH clusters. For doing that we need to set a threshold.
Tweaking threshold for clusters, plotting what we take into account and what no.
https://gspinaci.github.io/diary/10.10.2019.pdf

November 2019

Plot cumulative distribution normalized of DH pubs percentage

For each cluster create a positional vector of journals.
Euclidean distance and cosine similarity to cluster them.
https://gspinaci.github.io/diary/11.01.2019.pdf

Leiden clustering over euclidean distance on crossref. We get 13 clusters.
https://gspinaci.github.io/diary/11.15.2019.pdf

Use the hierarchical clustering gettin 13 clusters to see the resulting value threshold.
https://gspinaci.github.io/diary/11.22.2019.pdf

December 2019

Continue to tweak DH threshold to categorize clusters as DH clusters.
https://gspinaci.github.io/diary/12.20.2019.pdf

We decided to change the way to retrieve clusters DH. Now on we use only percentages, no threshold anymore

We should find a measure to see DH → DH/DH → AH citation ratio

Use ERIH Disciplines to categorize AH clusters and their major disciplines

Decided how the data should be displayed

February 2020

For each cluster we have the % of DH, AH, DH-AH, and other (nodes file)
First in separate files
https://gspinaci.github.io/diary/02.14.2020.pdf

MS Academic is still under evaluation

March 2020

Nodes file in the same cluster
Now in the same file
https://gspinaci.github.io/diary/03.06.2020.pdf

Update list of DH Journals integrating 3rd iteration (disambiguation done as before ie choose the prevalent by the three of us)
Update the VosViewer maps and update clusters' percentages

Starting to write csv file containing citation percentages (edges file) not finished yet