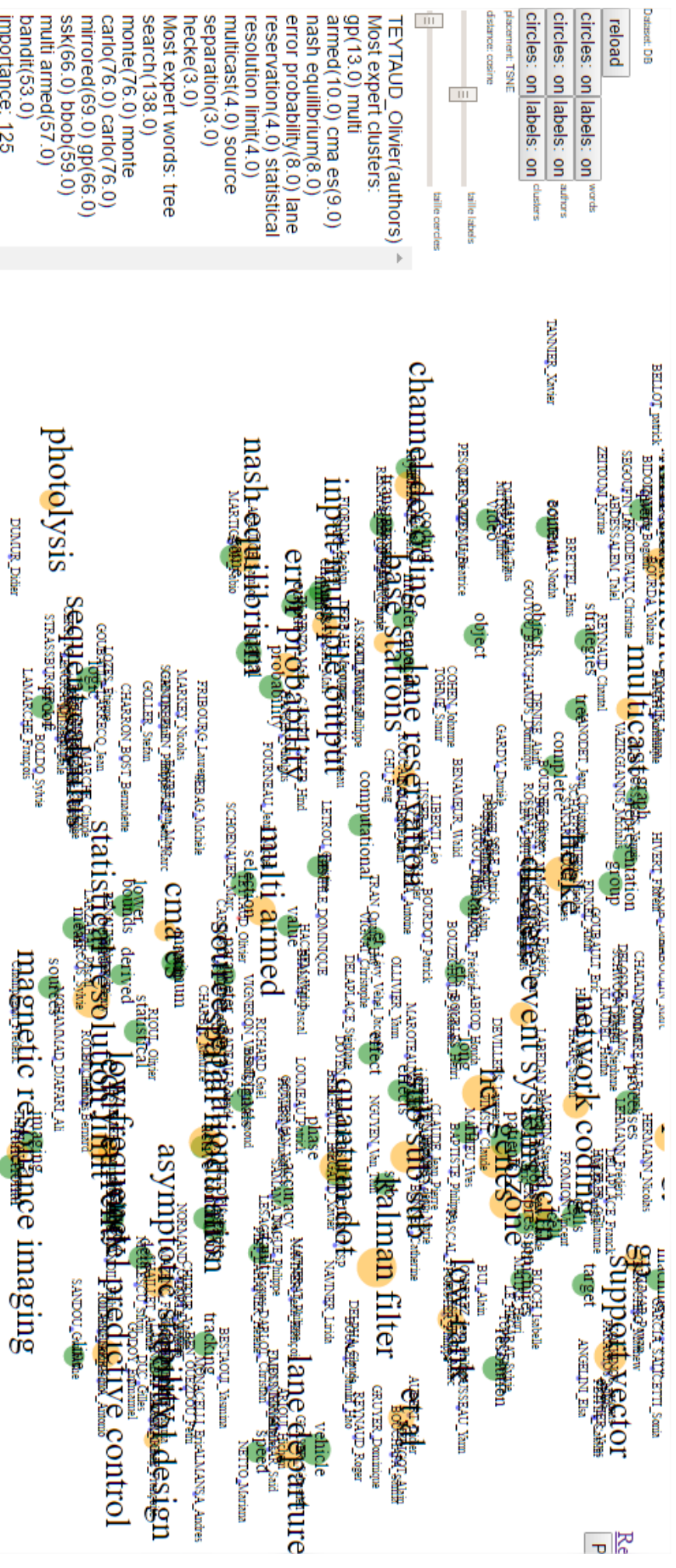


# GSPS

# Global Scientific Positioning System for CDS

Philippe Caillou (URPSud, INRIA TAO), Jean-Daniel Fekete (INRIA AVIZ)



# Goal: Visualize a scientific map using a scientific article database

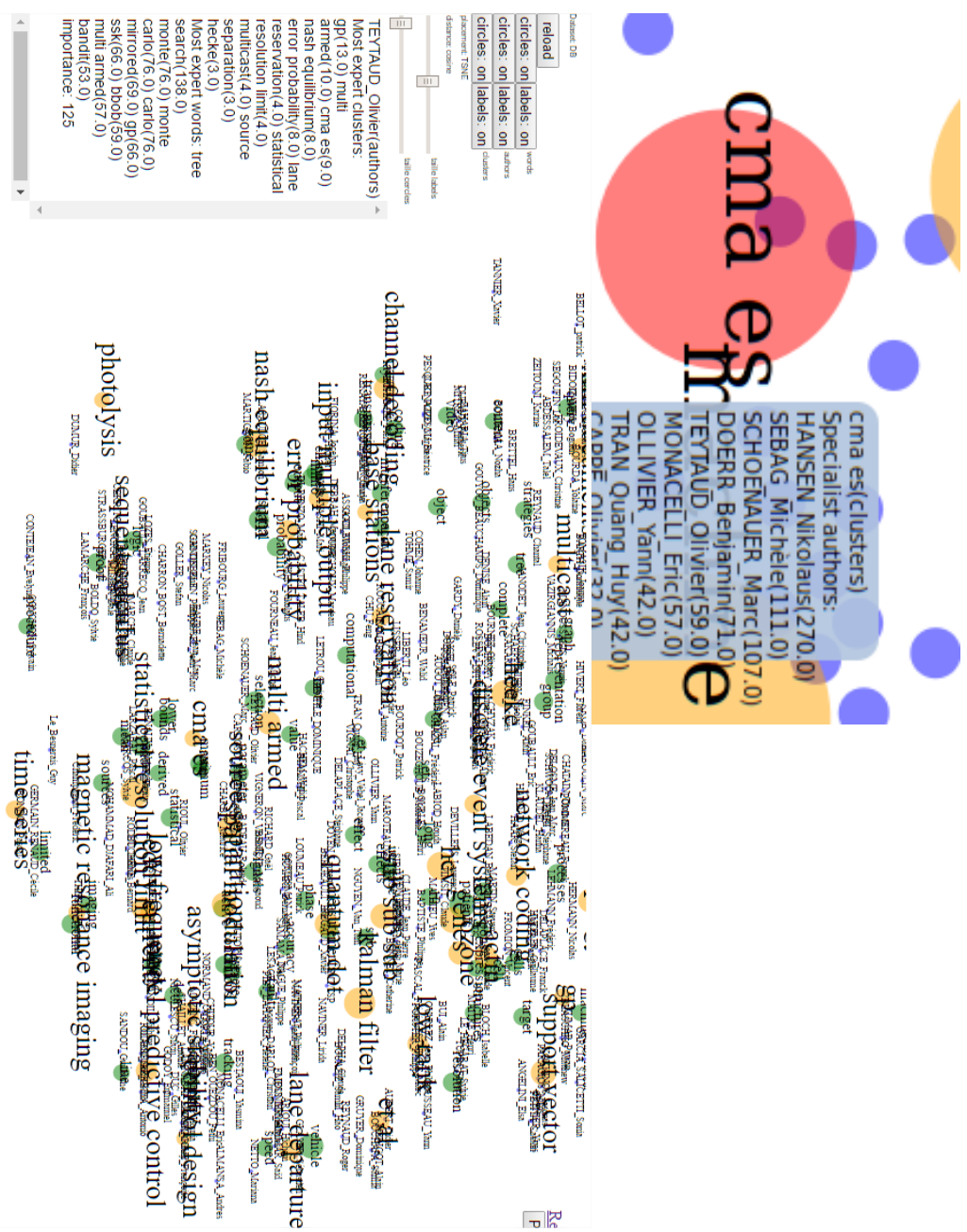
## Use Cases

- **Manager:** Get an overview of the research topics, their repartition and their evolution
- **Scientist:** Find colleagues on specific areas and their related topics

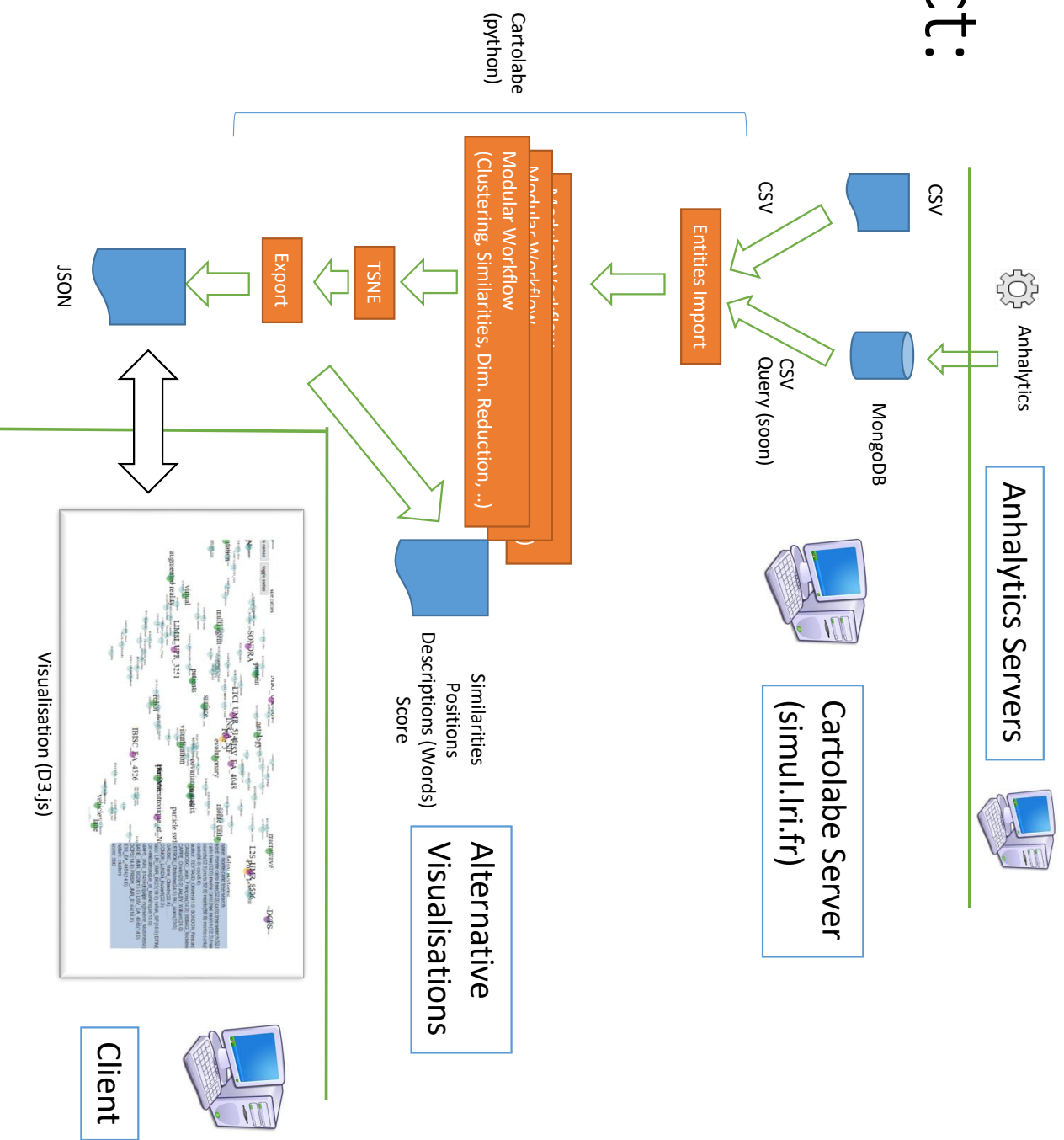
- **PC Chair:** Find someone for a review or a scientific committee

## Interest from CDS, INRIA, CNRS, DigiCosme

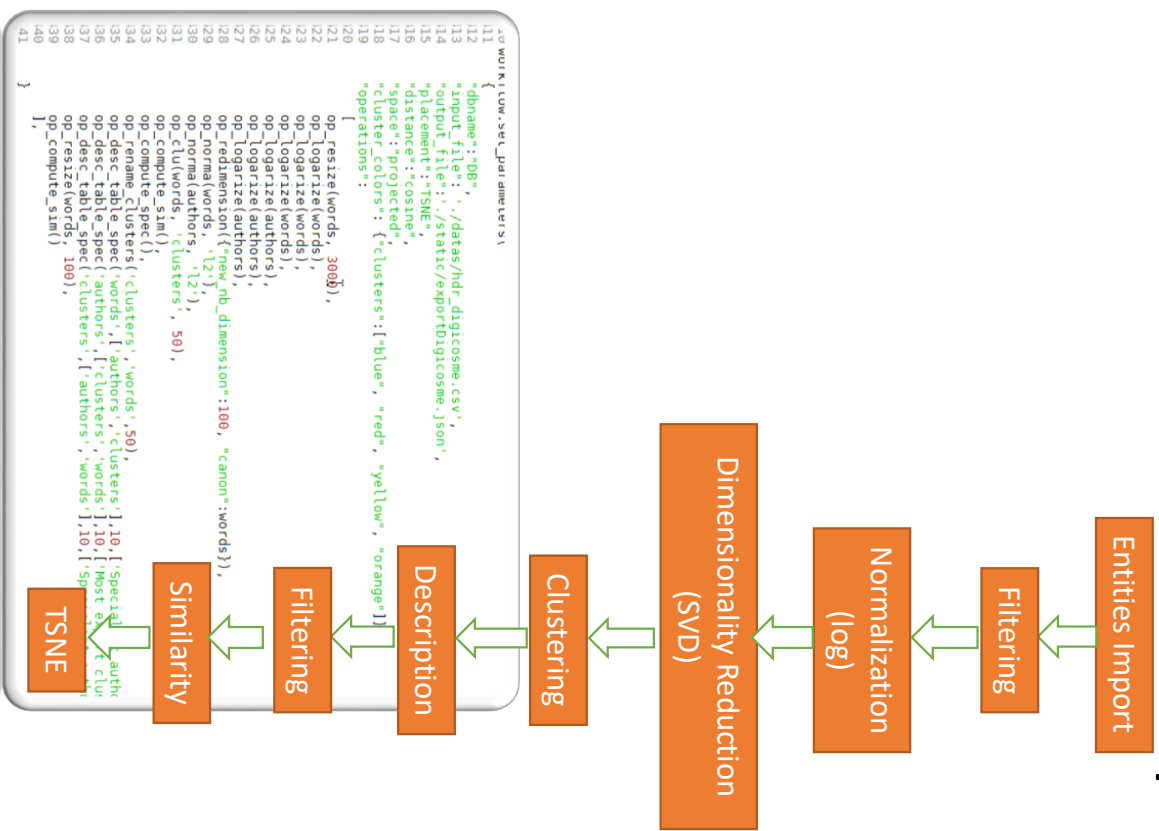
- ADT INRIA Cartolabe
- 2 years engineer position
- Started February 2016 (Felix Louistisserand)



# Cartolabe project: Overview



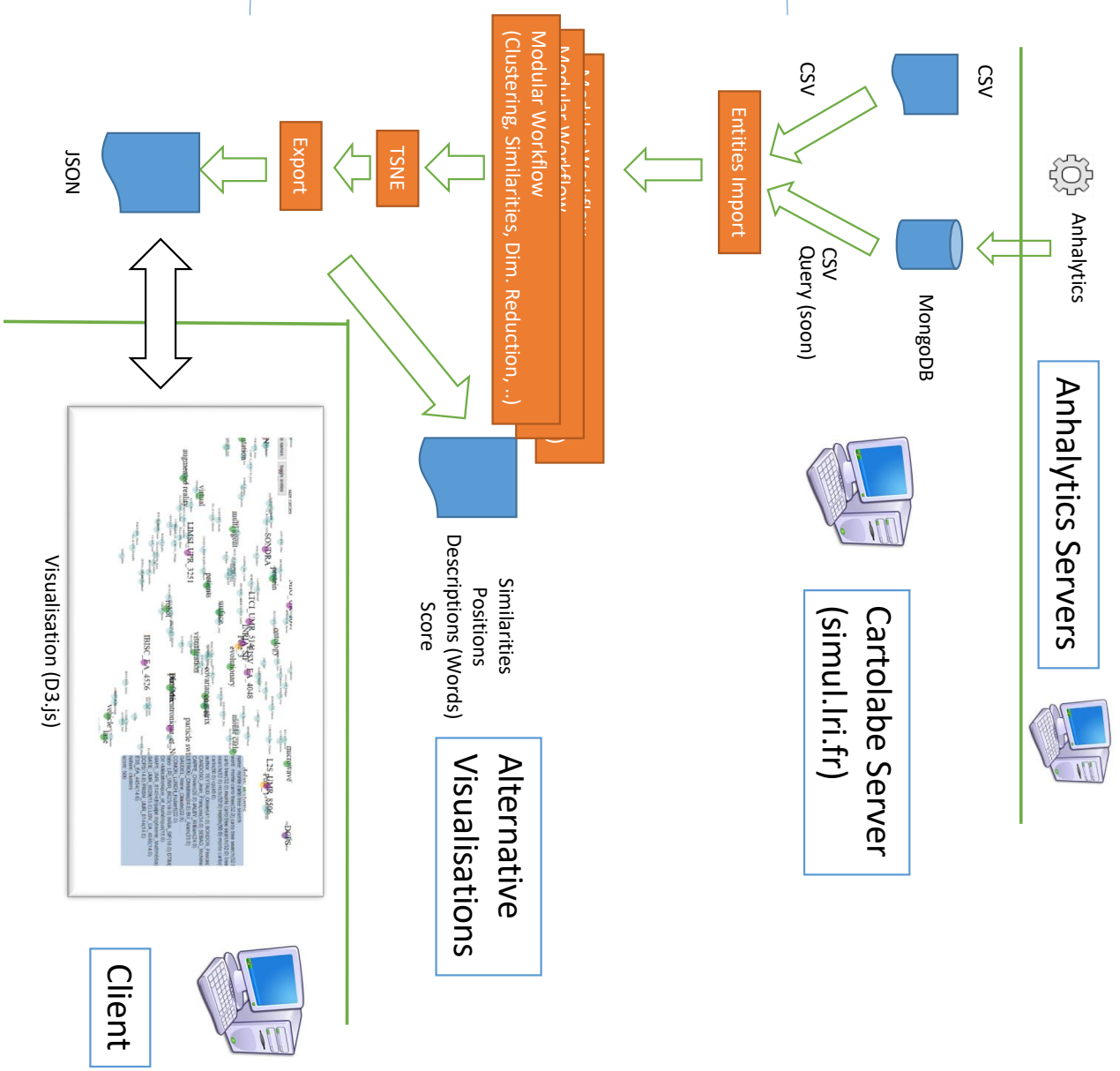
# Workflow example



```

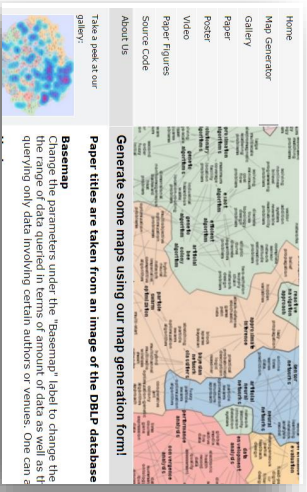
11 }
12 {
13   "dbname": "DB",
14   "input_file": "/data/hdr_digdigisme.csv",
15   "output_file": "/static/exportdigdigisme.json",
16   "placement": "TSNE",
17   "distance": "cosine",
18   "space": "projected",
19   "cluster_colors": { "clusters": [ "blue", "red", "yellow", "orange" ] },
20   "operations": [
21     op_resize(words, 3000),
22     op_logarize(words),
23     op_logarize(words),
24     op_logarize(words),
25     op_logarize(authors),
26     op_logarize(authors),
27     op_logarize(authors),
28     op_dimensional(new_nb_dimension:100, "canon":words),
29     op_norma(words, 12),
30     op_norma(authors, 12),
31     op_clu(words, clusters, 50),
32     op_compute_sam(),
33     op_compute_spec(),
34     op_rename_clusters('clusters', 'words', 50),
35     op_desc_table_spec('words', ['authors', 'clusters'], 10, ['Special', 'auth
36     op_desc_table_spec('authors', ['clusters', 'words'], 10, ['Most ex
37     op_desc_table_spec('clusters', ['authors', 'words'], 10, ['Sp
38     op_resize(words, 100),
39     op_compute_sam()
40 ],
41 }
  
```

Cartolabe (python)

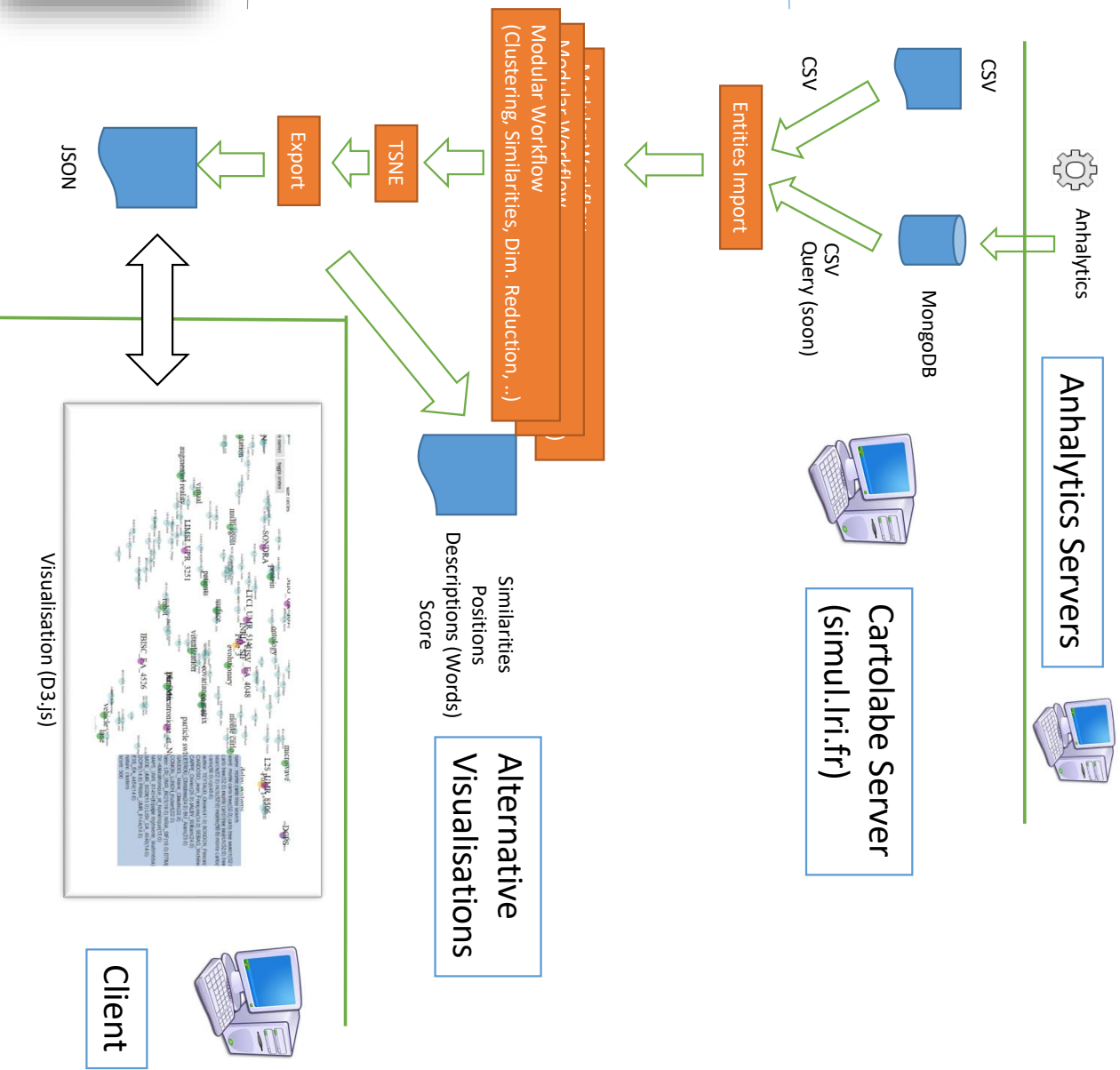
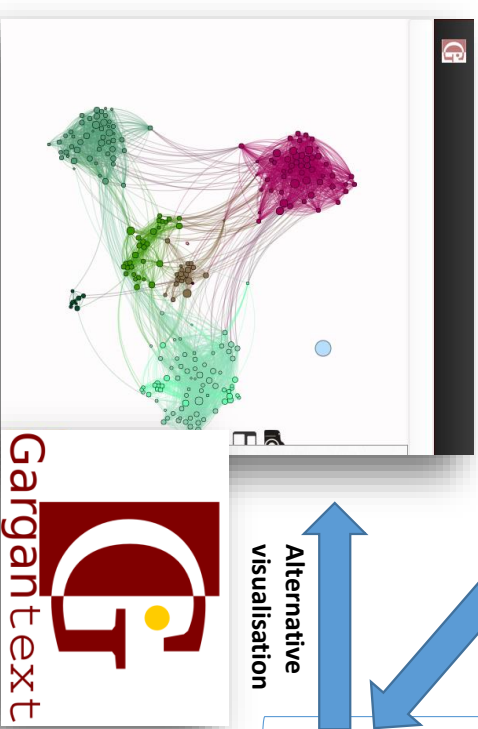


# Cartolabe alternatives

<http://mocs.cs.arizona.edu/>



<http://gargantext.org/>



Compare  
For specific  
people

Alternative  
Visualisation

# Cartolabe alpha

← → ↻ <https://cartolabel.fr>

- Applications
- ★ Bookmarks
- Index EDT
- sas-u-psud.fr

Welcome to Cartolabe (beta version) !

Here you can visualize the following datasets :

our research team :



a labex (university scale)



HAL, reformatted by AnF



Thanks for your comment

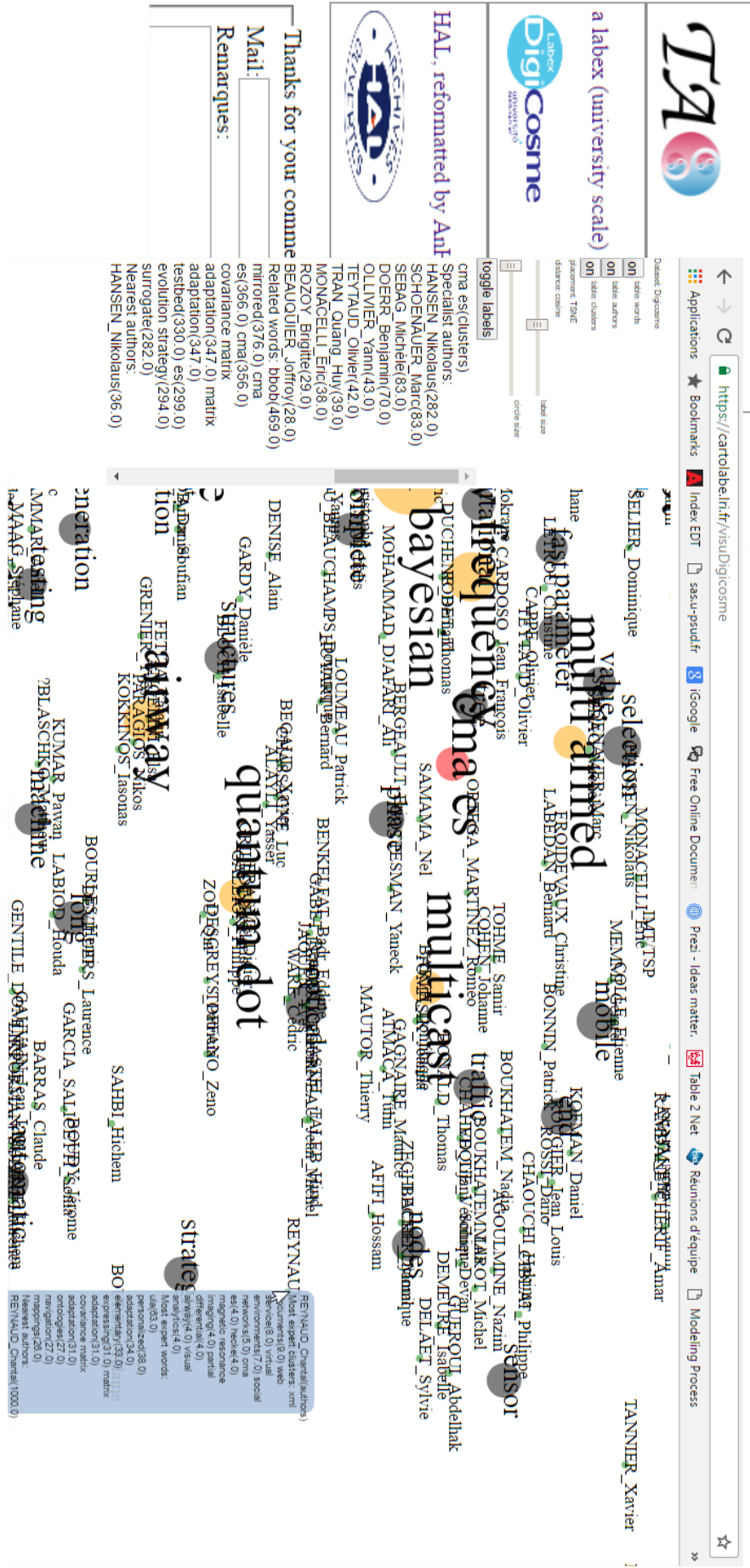
Mail:

Remarques:

Dataset Digicosme

title words  
 title authors  
 title clusters  
 placement\_TSNL  
 distance centers  
 label size  
 order size

toggle labels  
 cma es(clusters)  
 Specialist authors:  
 HANSEN\_Nikolaus(282.0)  
 SCHOENAUER\_Marc(83.0)  
 SEBAG\_Michiel(83.0)  
 DOERR\_Benjamin(70.0)  
 OLLIVIER\_Yann(43.0)  
 TETAUD\_Olivier(42.0)  
 TRAN\_Quang\_Huy(39.0)  
 MONACELLI\_Eric(38.0)  
 ROZOY\_Brightie(29.0)  
 BEAUOULIER\_Joffroy(28.0)  
 Related words: hbob(469.0)  
 mirrored(376.0) cma  
 es(366.0) cma(366.0)  
 covariance matrix  
 adaptation(347.0) matrix  
 testbed(330.0) es(299.0)  
 evolution strategy(294.0)  
 surrogate(282.0)  
 Nearest authors:  
 HANSEN\_Nikolaus(36.0)



REYNAUD\_Chantal(authors)

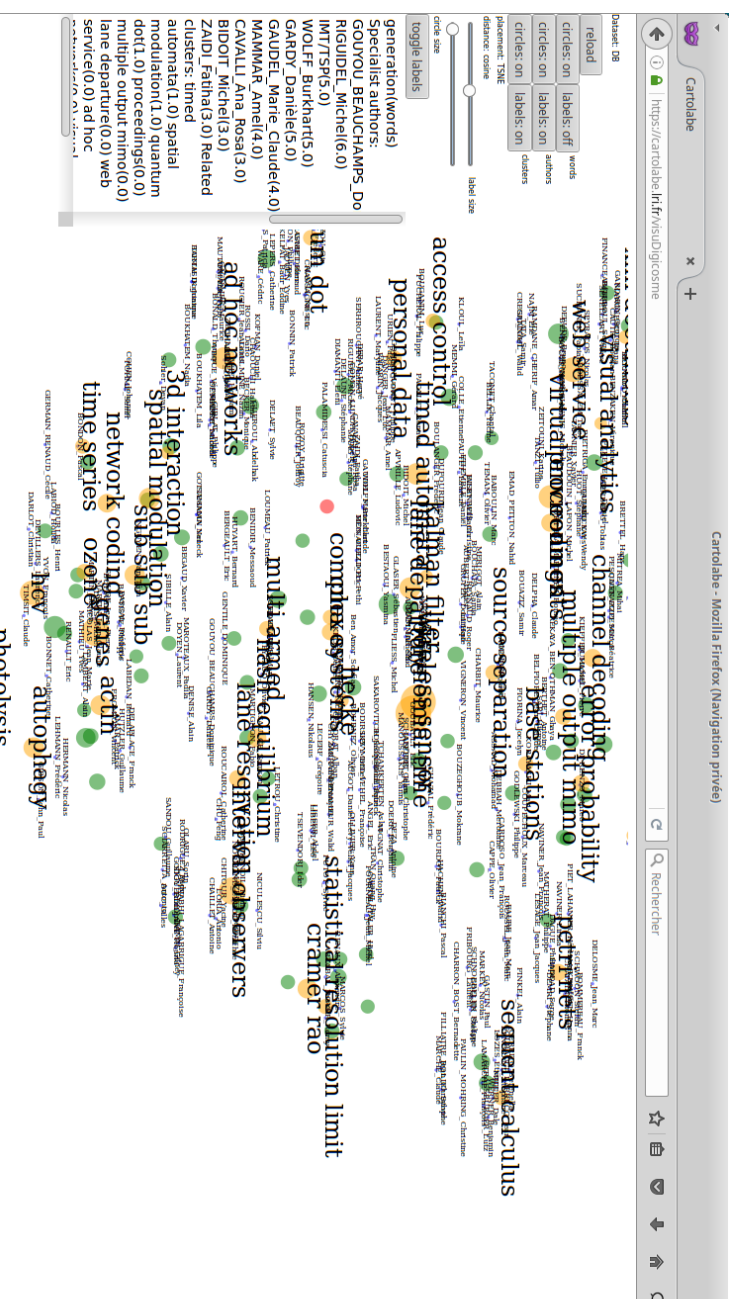
Not expert clusters: xml  
 journals(19.0) web  
 services(8.0) virtual  
 remote(14.0) social  
 es(4.0) neural(4.0)  
 magnetic resonance  
 imaging(4.0) partial  
 differential(4.0)  
 survey(4.0) visual  
 Most frequent words:  
 using(63.0)  
 personalized(38.0)  
 eLearning(34.0) http  
 expressing(31.0) http  
 adaptation(31.0)  
 ontologies(31.0)  
 adaptation(27.0)  
 navigation(27.0)  
 mpeg(26.0) os(8.0)  
 Nearest authors:  
 REYNAUD\_Chantal(1000.0)

# Cartolabe.Iri.fr

## Open Beta next week!

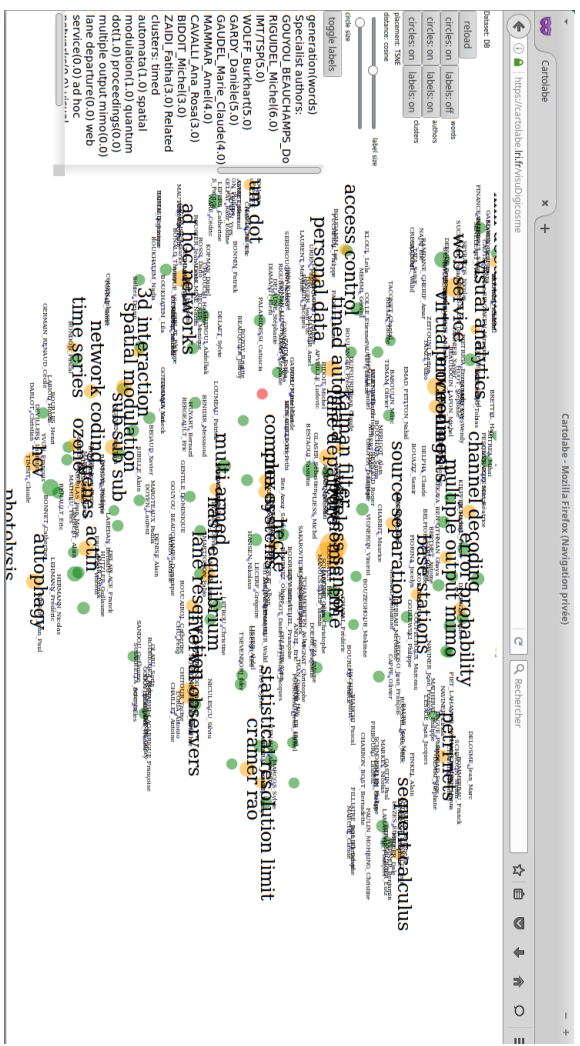
If interested: caillou@iri.fr

- Planned in the ADT/internships
  - Scale challenge (HAL database)
  - Multi-level / Progressive t-SNE
  - Dynamics/trajectories
  - Sub maps from queries
  - Alternative representations
  - Alternative bricks (LDA vs LSA)
- A lot is NOT planned



# Post-Doc objective 1: Assess the quality of the result

- Goal: hyperparameters optimization
- With what data?
  - Self declared keywords
  - Coauthorships/cocitations networks
  - Laboratories/teams dispersions



Which one is the best? Which cluster names?

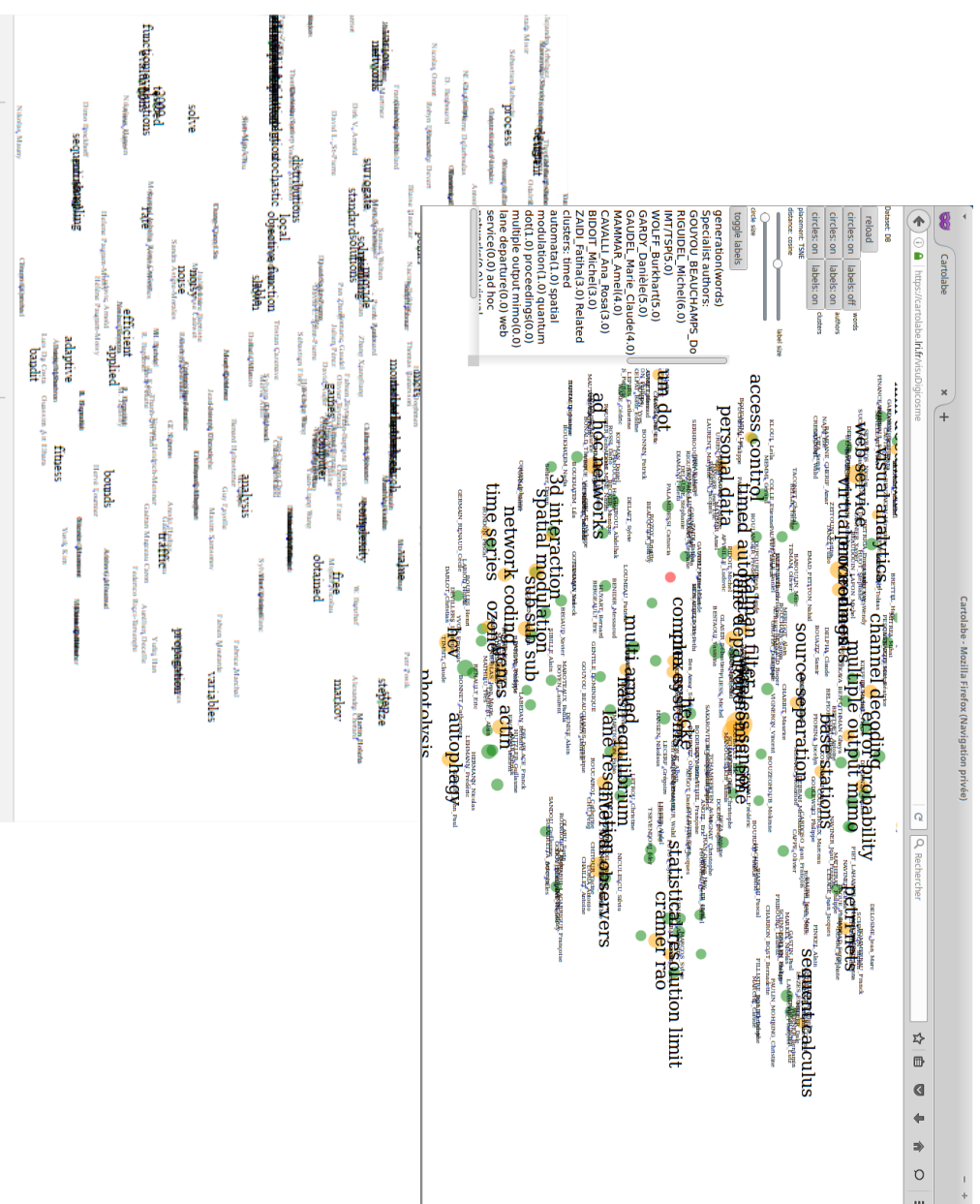




# Post-Doc objective 2: Different maps for different users

The journalist and the scientist do not want the same maps nor the same words or even the same entities!

- Similarity/preference learning from the expert/users feedback
- Users clustering to define user profiles/use cases
- Hyperparameter auto-tuning for each user profile



# At the end...

## A useful tool for CDS and the community

- Website to explore CDS
- Flexible open-source tool for researchers
- Maps adapted to various user profiles
- RAMP on CDS members cartography?

