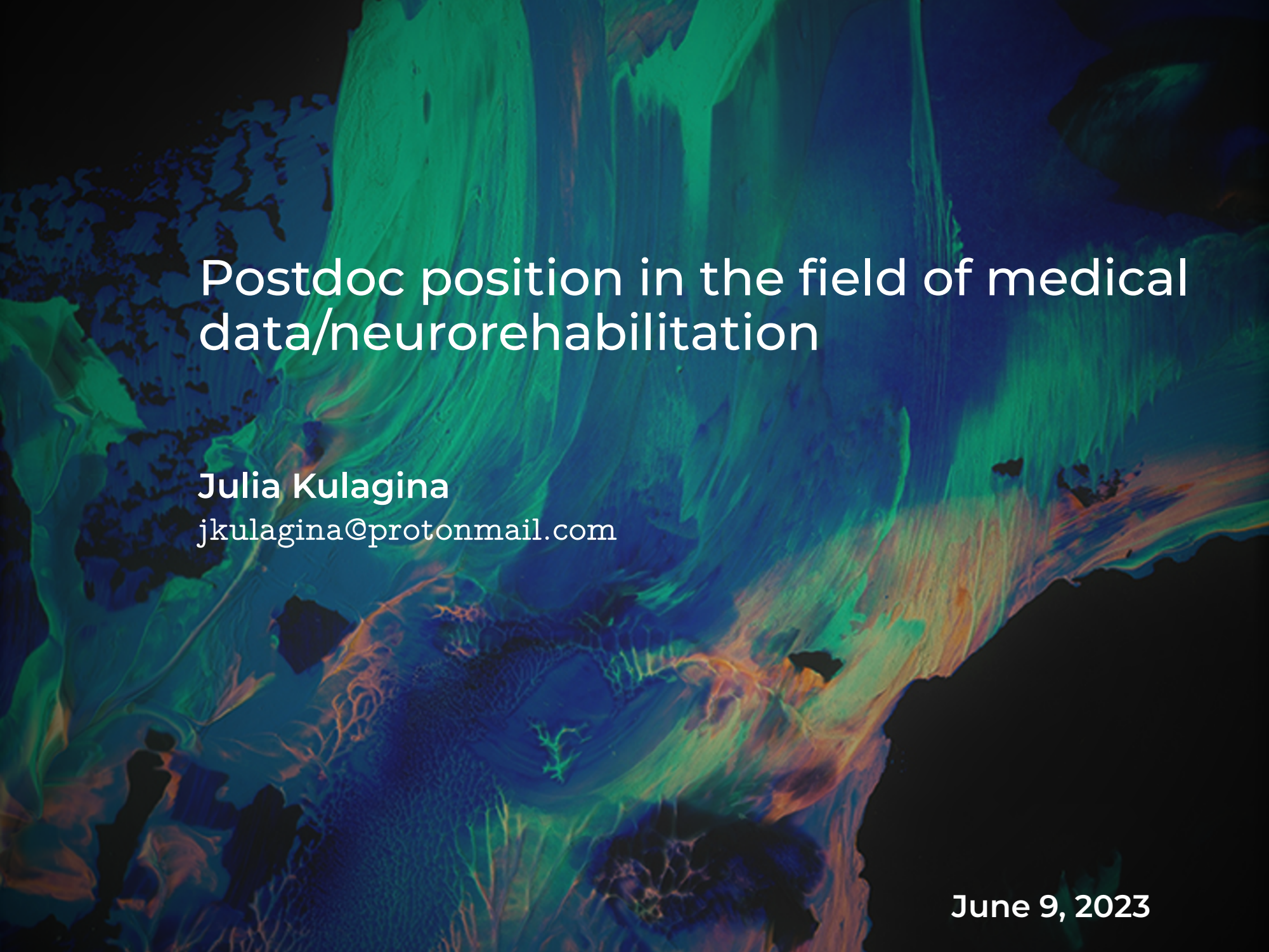


Postdoc position in the field of medical data/neurorehabilitation

Julia Kulagina
jkulagina@protonmail.com

June 9, 2023

Contents

1. Motivation
2. Experience/ Projects
3. PhD Thesis
4. Tools
   - 4.1 Technical Tools
   - 4.2 Data Pre-processing Tools
   - 4.3 Modeling Tools
   - 4.4 Visualization Tools
5. Interests
   - 5.1 Knowledge Bases
   - 5.2 Inspiration
6. References

Motivation

1. RESC /RELab mission: enhance quality of life for people with physical disabilities by promoting the establishment of a holistic rehabilitation approach, facilitating research and transfer of knowledge, fostering an interdisciplinary network; using robotics, wearable sensor technologies and non-invasive neuroimaging to explore, assess and restore sensorimotor function;
2. knowledge integration: interdisciplinary network, state of the art ML/DL tools for effective decision-making through collaboration between academia circles, state-officials, industry, community;
3. Personal development: real life data, programming, data architecture, creativity/ flexibility, new skills (new models/ architectures implementation, dashboards, visApps, KGs), finding patterns/ interconnections (knowledge retrieval through ML tools);
4. Further opportunities: experience (know-how) transfer.

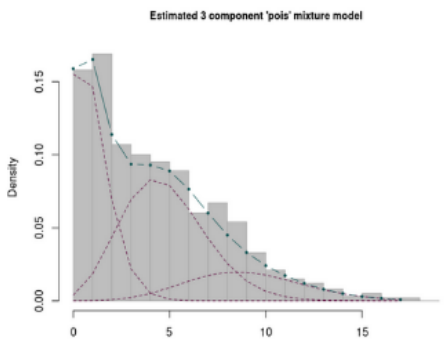
Experience/ Projects

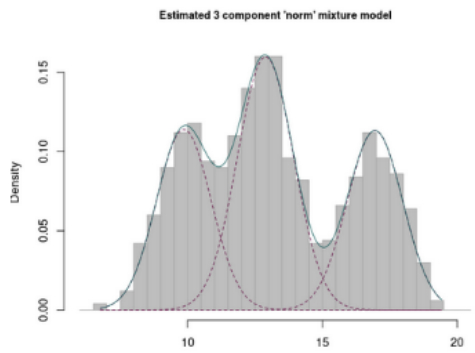
- *Work:*:
  - Relational DBs, DWH - mostly as a user (SQL);
  - Data processing, modelling/ forecasting, visualization;
  - Automation of business processes;
  - Pilot projects management (project design, liaison with IT, legal and accounting dep. & communication with external agents, results analysis).
- *Academia:*:
  - Master Thesis: Neural ODEs (*Matlab, Python*)
  - Stats Lab: MDMA Effects Statistical Analysis (*R*)
  - PhD: Mixture Models (Complexity Estimation Techniques, Applications)
  - Teaching: Applied Time Series, Stochastic Simulation, Student Seminar

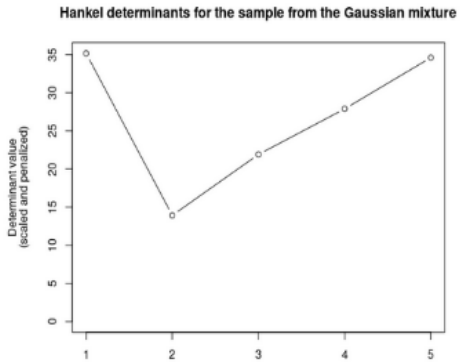
PhD Thesis

- Species richness estimation (a shape-constraint-based estimator)
- Complexity estimation: `R` package `mixComp`, ML methods

```r
# apply papamHankel.scaled to datMix objects:
set.seed(1)
pois_sca_pen <- paramHankel.scaled(pois.dM)
norm_sca_pen <- paramHankel.scaled(normLoc.dM)
# plot the results for both mixtures:
par(mar=c(5, 5, 1, 1))
plot(pois_sca_pen,)
plot(norm_sca_pen)
```





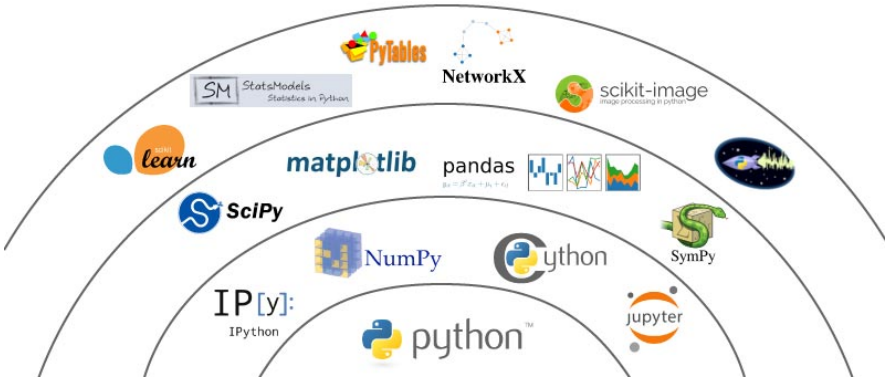


- Data processing, modelling, visualization, dataset-building
- Clustering, structure retrieval, lyrics generation

Technical Tools

1. Python: pandas, sklearn, scipy, sympy, tensorflow, keras, plotly, seaborn, kglab ...
2. Git + GitHub
3. R
4. other: SQL, HTML, LaTex, LibreOffice

Data Pre-processing Tools

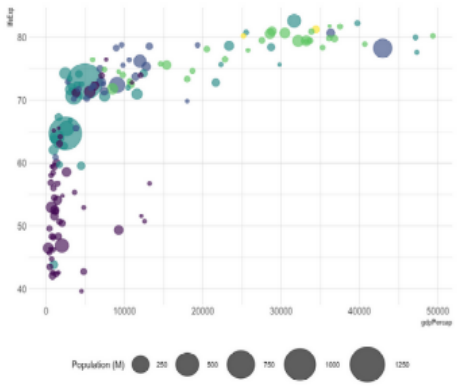


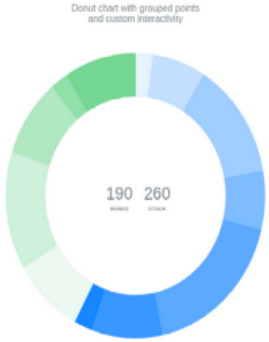
1. Exploratory DA (sklearn, matplotlib, plotly, pandas)
2. Duplicates, errors, outliers (anomaly detection), MV's (sklearn, pandas)
3. Categorical, binary, temporal, geographical, textual variables: encoding (tensorflow, numpy)
4. Normalization, scaling, binning (numpy, tensorflow, keras)
5. Feature engineering (KGs)
6. Sample split (train, test, val) (sklearn, tensorflow, keras)

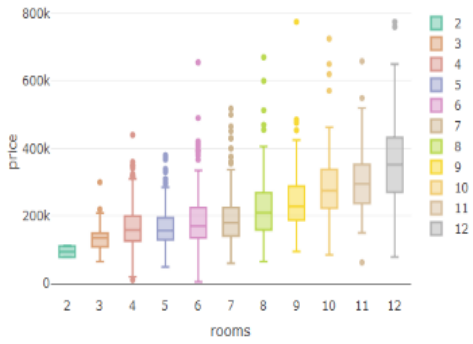
Modeling Tools

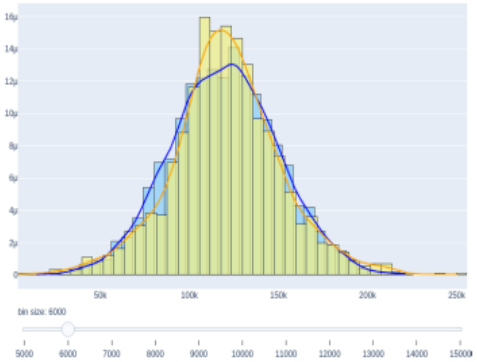
1. Regression (simple, multiple)
   - Linear (robust, regularized - Lasso, Ridge)
   - Non-linear (K-Nearest Neighbor, Polynomial regression, Decision trees and random forests (bagging, boosting, Support Vector Regression (RBF-kernel), Neural networks)
2. Classification (Logistic Regression, Decision trees and random forests, KNN, SVM, Multi-Layer Perceptron, ConvNets,)
3. Clustering (K-means, Density-based Spatial Clustering (DBSCAN), Mixture models)
4. Dim-reduction (Principal Component Analysis, Linear Discriminant Analysis, Non-negative matrix factorization)
5. Time Series ((S)ARIMA, (G)ARCH, DL: RNNs, LSTMs)

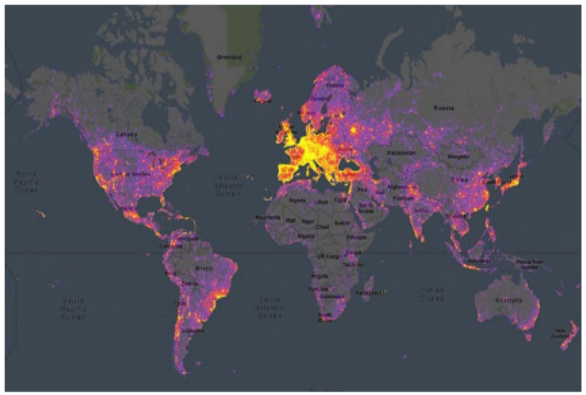
Visualization Tools

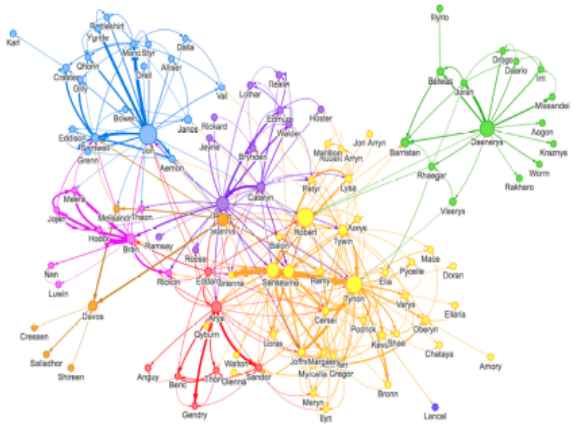


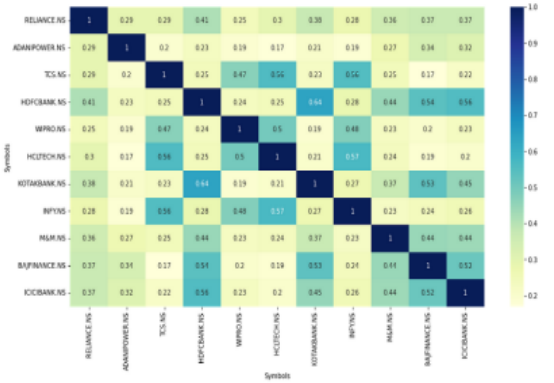












Interests

- VisApp tools (Python Dash, Bokeh)

- Generative AI (GANs, cGANs, transformers)
- Knowledge Graphs (SPARQL, OWL, RDF/ RDFs)

- Data sharing (Git, Zenodo, Kaggle)

Knowledge Graph
Defenition and Uses (General)

Def:
A **KG** showcases key structural aspects: concepts entities attributes typologies aspects
as well as supports direct and inferential browsing of the structure.

Key Properties/ Uses:

1. Learning about individual concepts and entities
2. Discovering related concepts and entities
3. Understanding the structure and typologies
4. Tracing conceptual lineages
5. Exploring inferences based on the logical assertions in the graph
6. Parallel access to structured and semantic search
7. Information source for both people and machines

Knowledge Graph + ML I

ML on KG to answer neuro-rehab relevant questions

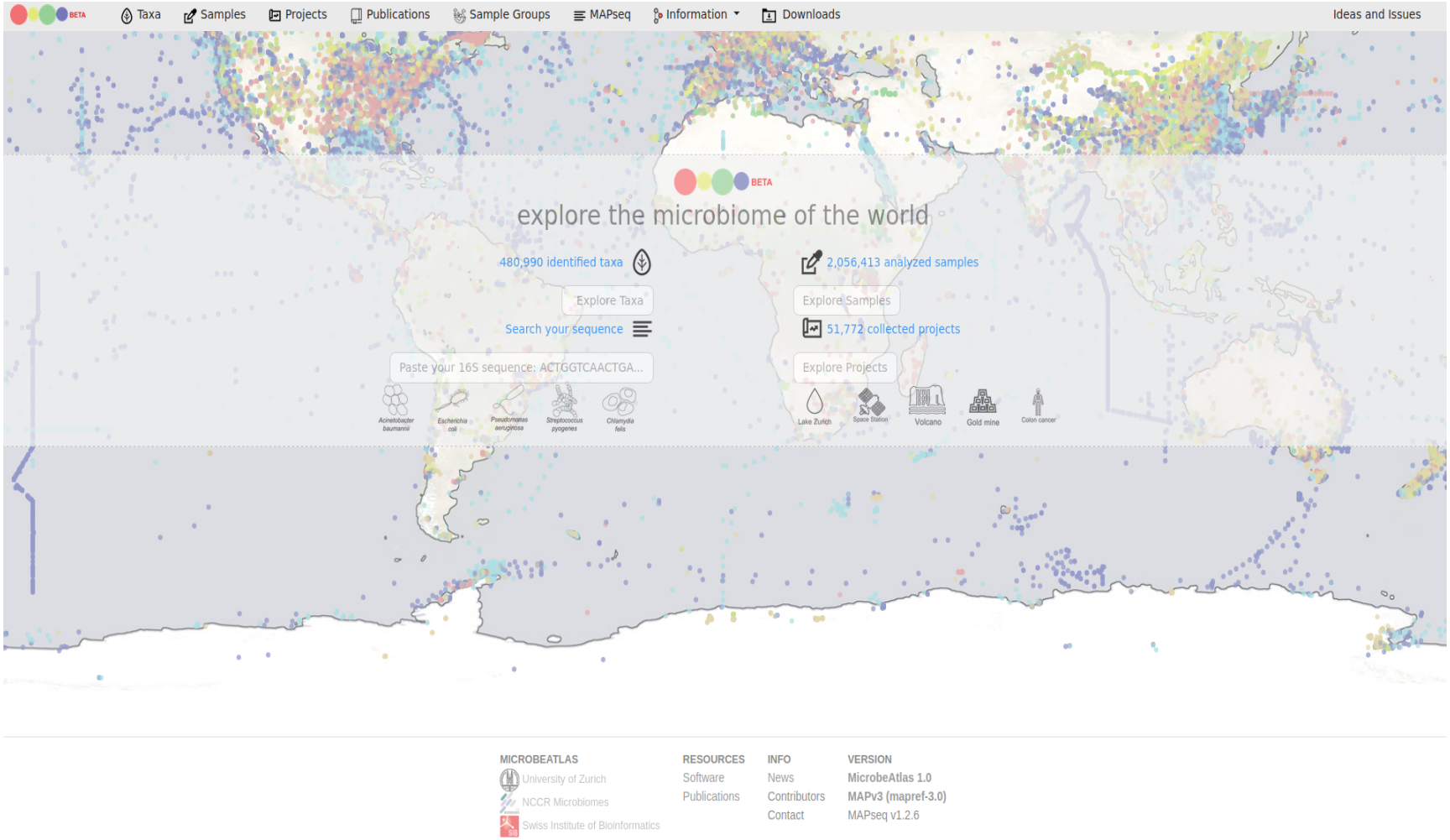
1. Which groups of people are most prone to strokes?
2. Which neuropsychological after-stroke consequences patients might have that cannot be identified at the moment?
3. Which rehab method will be most effective in a given situation (patient psychological, emotional and physical state, medical history, etc.)?
4. Which prevention methods might be most effective in a given case?
5. How long might it take to restore certain body functions given a certain rehab program / aids?
6. Which health conditions might slow down the rehab (interaction effects) and how to take them into account?
7. Which new methods (rehab programs, robotics, wearable technologies) can be developed?

Knowledge Graph + ML II
ML on KG to answer neuro-rehab relevant questions

by means of:

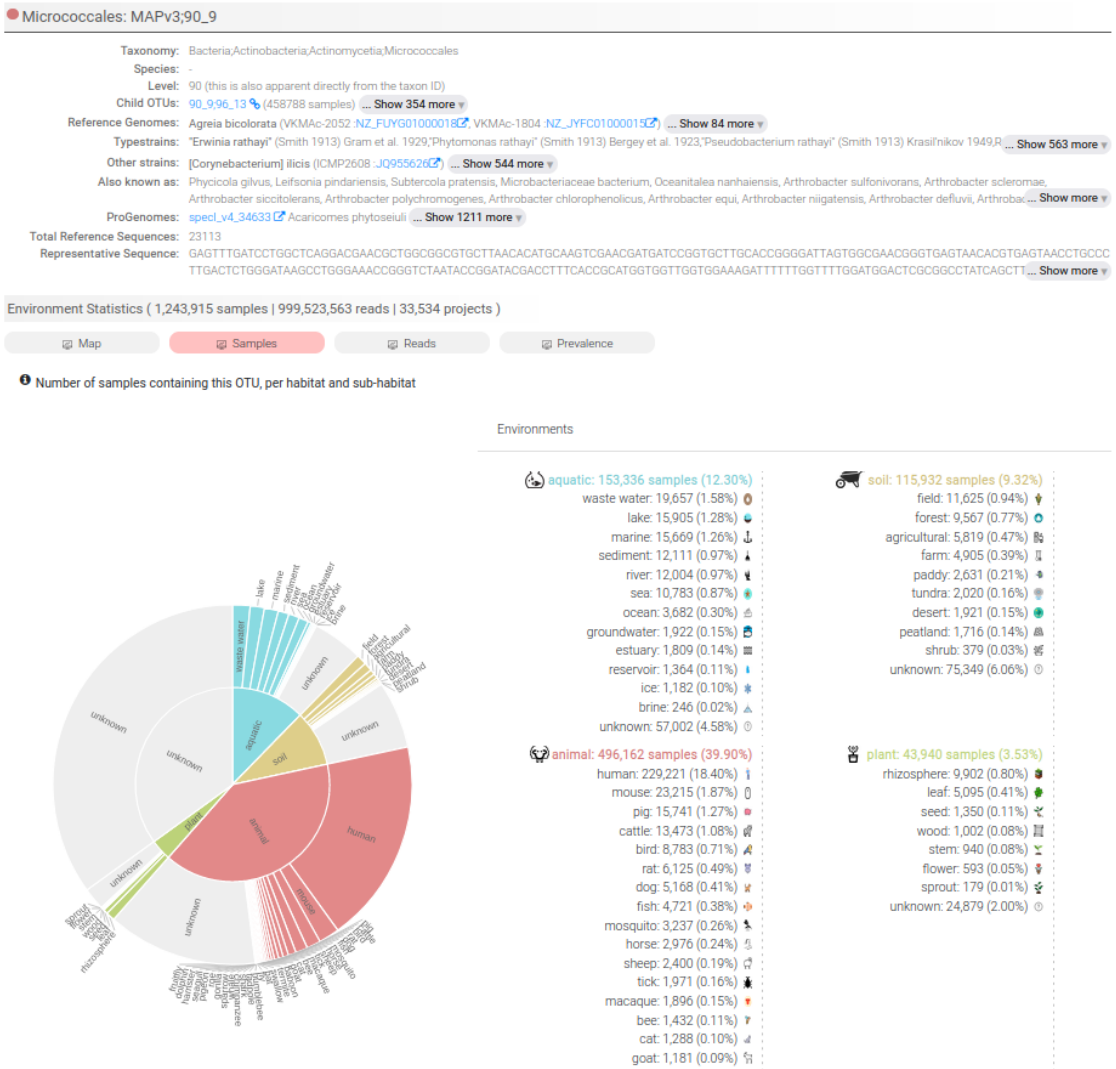
- tracking changes in relations (edges) between entities/ concepts
- modeling behavioral / physical results given patient input data
- accounting for multiple interaction effects due to patient health condition
- filling in the gaps in the nodes (entities) and edges (relations)
- generating new possible rehab programs, assistive technologies/ equipment
- forecasting
- aid technology reliability analysis, error monitoring and forecasting
- RecSys on top of KG, etc.

MICROBEATLAS KG (Motivation) 1
Ideas



Source: https://microbeatlas.org/index.html?action=taxa

MICROBEATLAS KG (Motivation) 2
Ideas



Source: https://microbeatlas.org/index.html?action=taxa

References

1. Tom Heath and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.
2. USE CASE: Browsing the KBpedia Knowledge Graph: https://kbpedia.org/use-cases/browse-the-knowledge-graph/
3. MICROBEATLAS https://microbeatlas.org/index.html?action=taxa
4. MICROBEATLAS Git Repo https://github.com/meringlab/MicrobeAtlasWebsite