

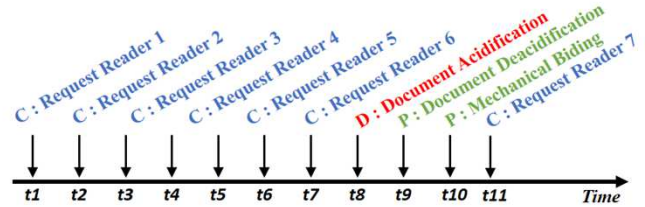
AI and Data to Improve Conservation Management

1 – Why DALGOCOL ?

- **Goal** : a decision helper software able to predict documents' physical condition without having to see them one by one to help librarians to identify collections in need for care treatments
- Dealing with 40 million documents, described in more than 15 million bibliographic records
- Analysing the data at the BnF
- Different tries to have a representation of the documents conservation history
- Predicting the documents' physical condition
 - Proposing an ontology to represent the conservation events
 - Proposing a new similarity measure
 - Clustering of the similar conservation histories

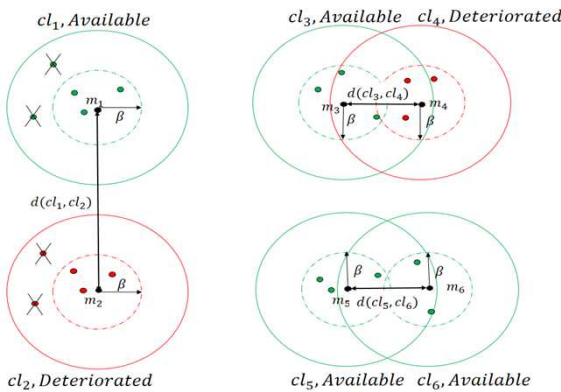
2 - Conservation History

- Relevant events to characterize the history :
 - Detected degradations
 - Conservation-restoration processes
 - Communication to the readers
- Represented by a semantic trajectory (sequence of events) :



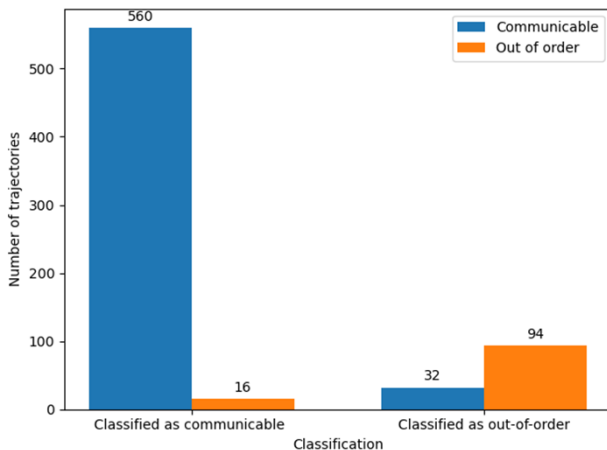
4 – Machine Learning to Predict the Physical condition

- Trajectory clustering
- Identifying trajectory patterns
- Defining prediction rules based on the patterns



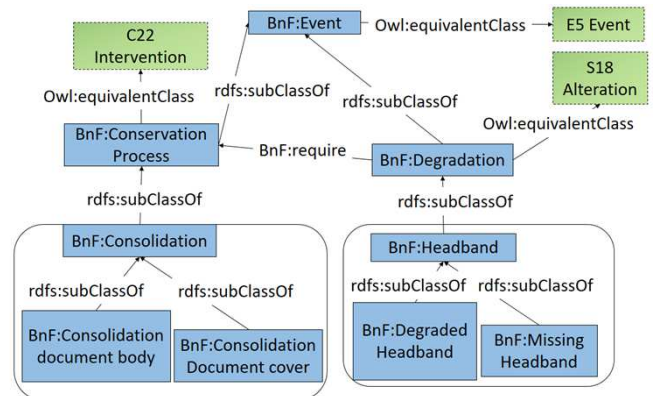
5 – Prediction Results

- Running a test on 702 documents, creating each trajectory and comparing it with the created clusters...
- ... the prediction algorithm classified 654 documents in the right communication conditions and 48 in the wrong ones.
- 7% of errors on the first try !



3 – “CRM BNF” Ontology

- Describe the concepts used in the BnF databases to describe the conservation history
- Used to resolve the terminological heterogeneity of different databases
- Used in the comparison of the conservation histories of the documents.



6 – What's next ?

- Run other tests as a proof on concept : especially on a set of documents that has been seen one by one by an other data project
- Work on data quality to enhance the prediction results
- Add more terms (and more data as well) to the « CRM BNF » Ontology to calculate new trajectories and identify new conservation patterns
- Add data now stored in other databases, such as documents' characteristics
- Add data extracted using text mining from work restoration files to databases
- Run other research projects to go on exploring what AI can do to improve conservation management !

References

- Zreik, A., Kedad, Z.: Matching and analysing conservation–restoration trajectories. Data Knowledge Engineering 139, 102015 (2022)

