Reacfin

Know-How to Risk

Contribution of Data Science to the Solvency 2 regulatory framework:

SFCR automated analysis

Aurélien Couloumy, ***Head of Data Science***
Aurelien.Couloumy@reacfin.com

Insurance Data Science Conference 2018
Cass Business School, University of London

I. Introduction – Data Science business cases and Solvency 2

- Use of Data Science could help facilitating S2 framework absorption by all parties

Model points + ML*

Risk profile + ML

Reduce calculation time within processes

Asset allocation + ML

Lapse rate + ML

Improve hypothesis and modeling

BSCR + ML

ORSA + ML

Give intuition and agility

Exploit results quickly and automate controls

ML* = Machine learning

I. Introduction – Context of the study

- Context:

- Constraints:

Understandable process

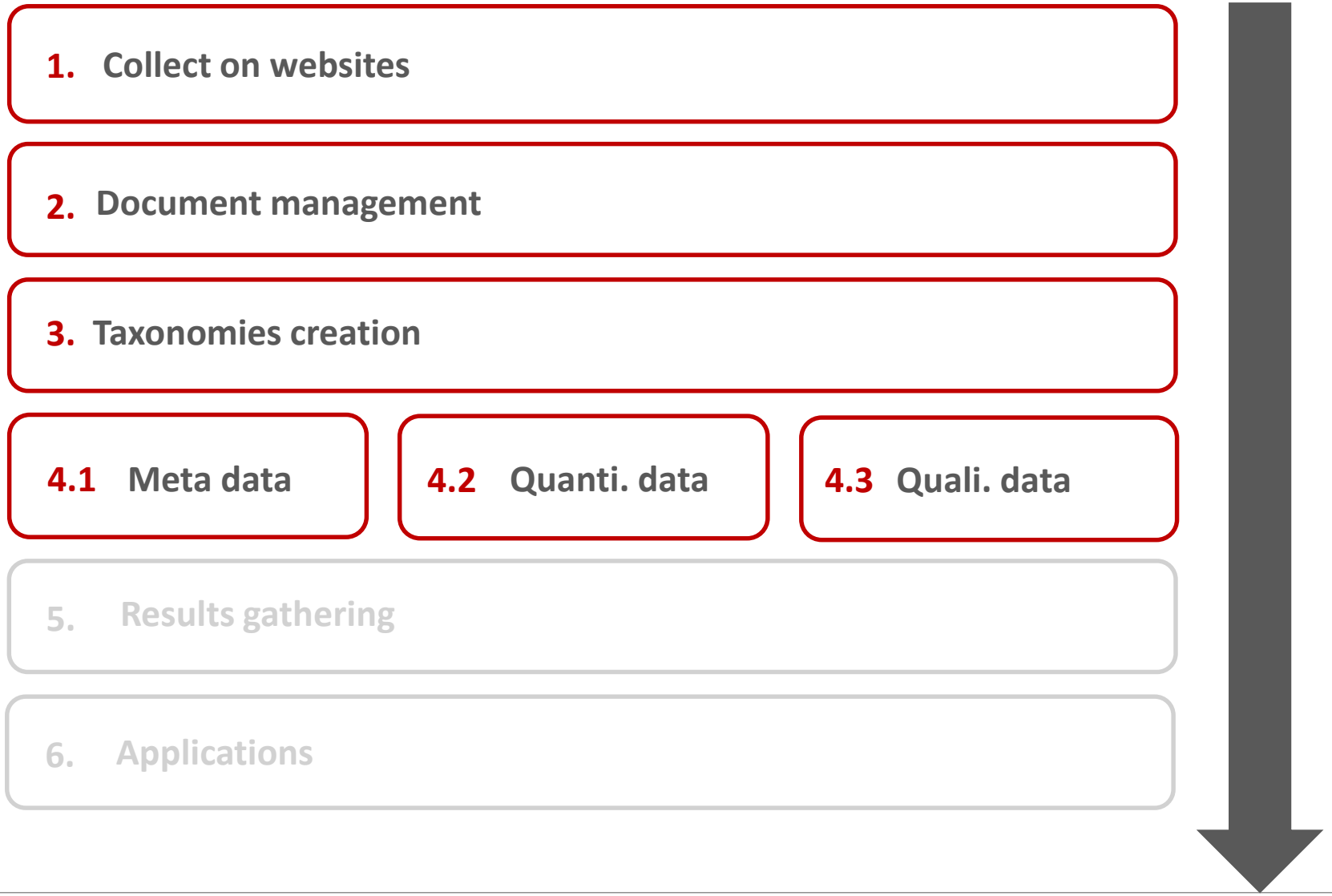
Light and open technology

Evolving methodology

- Goals:

1. **Analyse and control easily**
2. **Automate and save time**
3. **Create value for other applications**

II. Overall process – Structure



III. Results – SFCR context

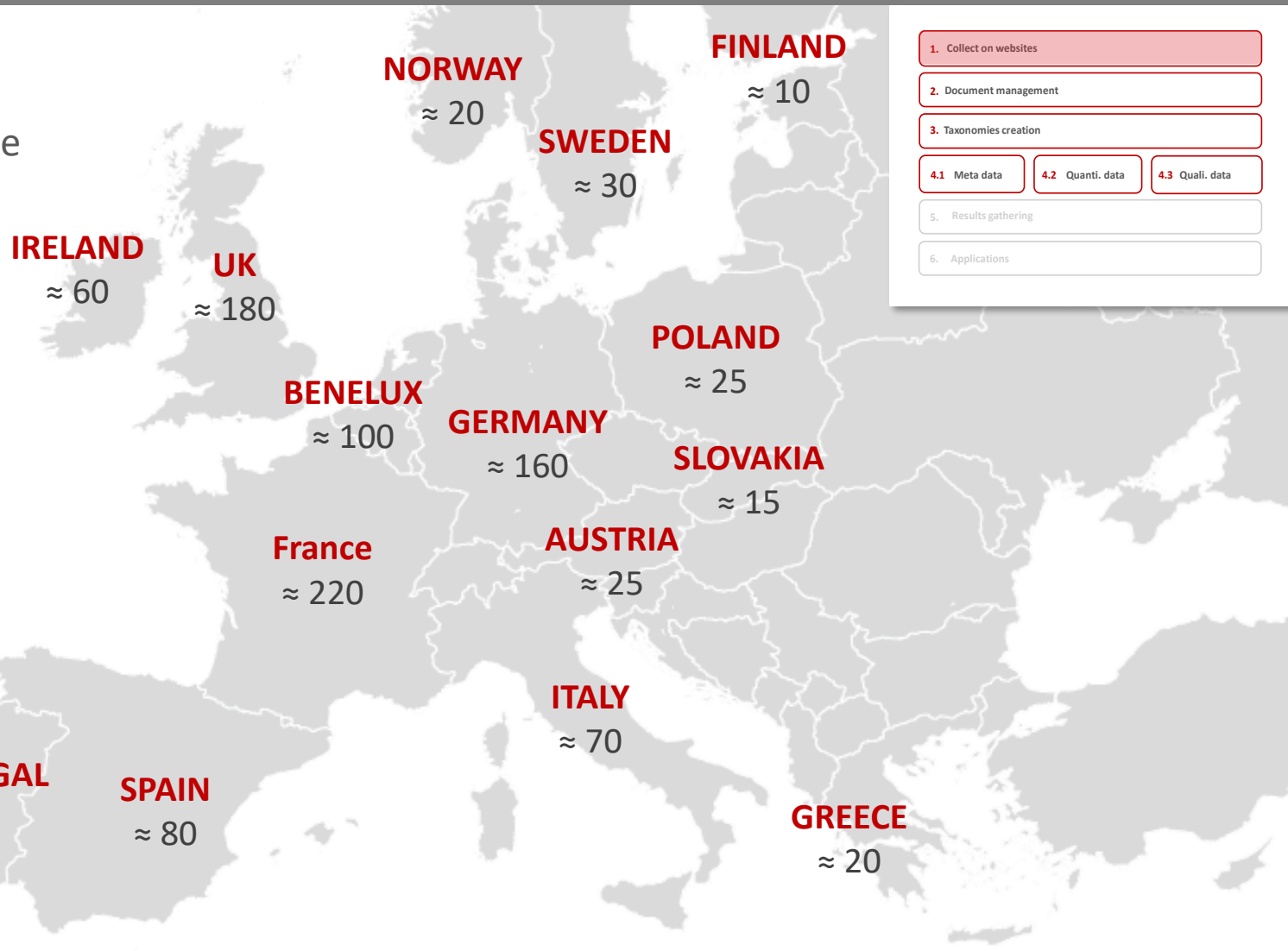
- Study realized on the former exercise (publication of 18th of May 2017)
  - Estimation of around 2400 reports available on the web

- Globally, in Europe, each company has played his role, but in an heterogeneous way
- A lot of consulting propose analysis reports without considering qualitative aspects

Application of the process to this SFCR context

III. Results – Collect on websites

- Extraction of the whole list of (re)insurers websites
- Around 1200 SFCR collected
- All countries are not available on the map
- Both Solo and Group reports



1. Collect on websites
2. Document management
3. Taxonomies creation
   - 4.1 Meta data
   - 4.2 Quanti. data
   - 4.3 Quali. data
5. Results gathering
6. Applications

Reacfin
Know-How to Risk

III. Results – Document management

25% to 28%

of rejected documents thanks to the **modification date**

5% to 7%

of rejected documents thanks to the **number of pages**

2% to 3%

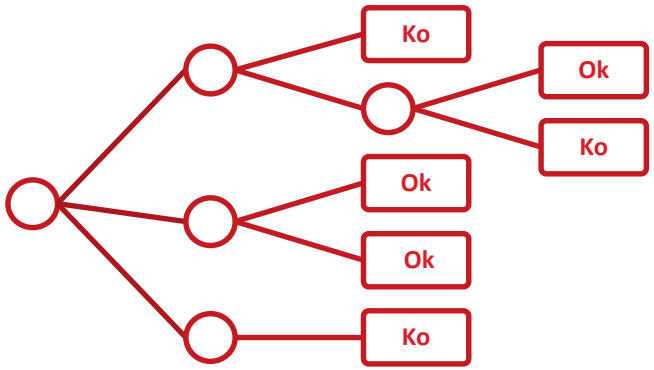
of rejected documents thanks to the **document size**

26% to 31%

of rejected document on average

1. Collect on websites
2. Document management
3. Taxonomies creation
   - 4.1 Meta data
   - 4.2 Quanti. data
   - 4.3 Quali. data
5. Results gathering
6. Applications

- We can also use supervised machine learning models to define a topic recognition model and to classify if documents are real SFCR

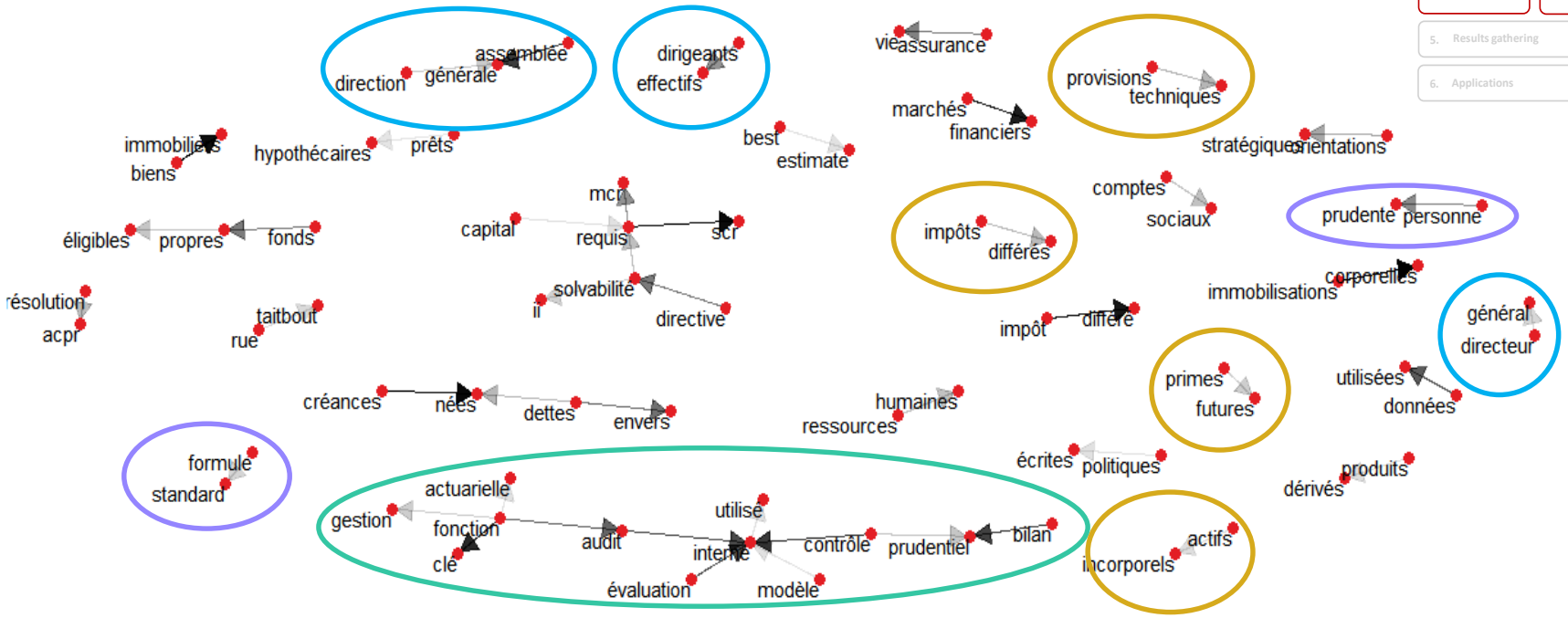


Reacfin
Know-How to Risk

III. Results – Taxonomies creation

- **Several automated approaches** to define the taxonomies and observe results

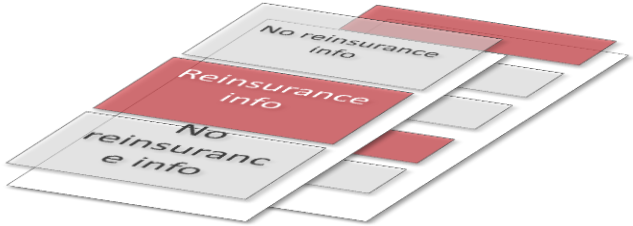
1. Collect on websites
2. Document management
3. Taxonomies creation
   - 4.1 Meta data
   - 4.2 Quanti. data
   - 4.3 Quali. data
5. Results gathering
6. Applications



- Definition of several characteristics of study: **LoB, Risks, Regulators, Outsourcer, Software, Countries, Modelling hypothesis, Reinsurance, Op. risk , Action Management**, etc.

III. Results – Taxonomies creation

- Definition **of supervised machine learning models** to study the topics / the architecture of document (page / paragraph level)



1. Collect on websites
2. Document management
3. Taxonomies creation
   - 4.1 Meta data
   - 4.2 Quanti. data
   - 4.3 Quali. data
5. Results gathering
6. Applications

- Definition of **metrics lists** (SCRs, turnover, number of emloyees, etc.) **and table templates** (QRTs) we would like to pick up in the documents

Les différentes composantes de ce que la Société considère comme des fonds propres éligibles sont déterminées selon les prescriptions de la réglementation Solvabilité 2.

Au 31 décembre 2016, les fonds propres éligibles sont de 60,2 millions d'euros répartis comme suit :

- Tier 1 non restreint après distribution de dividende anticipée correspondant au capital social composé d'actions ordinaires (51,3 millions d'euros) et de la réserve de réconciliation pour 8,9 millions d'euros, du fait de l'absence de dette subordonnée perpétuelle.
- Tier 1 restreint et Tier 2 : d'un montant nul du fait de l'absence de dette subordonnée.
- Tier 3 : actif net d'impôts différés dont le montant est nul car le montant net des impôts différés se trouve être au passif.

Aucun élément de fond propre auxiliaire n'a été identifié.

Les fonds propres éligibles de la Société, calculés selon la réglementation en vigueur, excèdent le minimum de capital requis (MCR). Le ratio de couverture du MCR s'élève à 538,32% au 31 décembre 2016 contre 406,56% l'année précédente.

(in miljoen euro)	Reële waarde (Solvabiliteit II)	Boekwaarde (Belgische boekhoudnormen)
Technische voorzieningen – Niet-Leven		
Technische voorzieningen – Niet-Leven (met uitzondering van Gezondheid)	**2.529**	**3.133**
'Best Estimate'	2.403	3.133
'Risk Margin'	126	-
Technische voorzieningen – Gezondheid (vergelijkbaar met Niet-Leven)	**561**	**1.201**
'Best Estimate'	559	1.201
'Risk Margin'	2	-
Technische voorzieningen – Leven (met uitzondering van Geïndexeerde en Rekeneenheden)		
Technische voorzieningen – Gezondheid (vergelijkbaar met Leven)	**3.153**	**2.025**
'Best Estimate'	3.045	2.025
'Risk Margin'	108	-
Technische voorzieningen – (met uitzondering van Gezondheid, Geïndexeerde e Rekeneenheden)	**29.416**	**25.197**
'Best Estimate'	29.227	25.197
'Risk Margin'	188	-
Technische voorzieningen – Geïndexeerde en Rekeneenheden	**2.188**	**2.215**
'Best Estimate'	2.168	2.215
'Risk Margin'	20	-

III. Results – Data

3 Data streams

1. Collect on websites
2. Document management
3. Taxonomies creation
4. 4.1 Meta data | 4.2 Quanti. data | 4.3 Quali. data
5. Results gathering
6. Applications

Meta data

- Company name ;
- URL ;
- Author;
- Creation/modification date;
- Page number;
- File size ;
- Encryptions level ;
- Language;
- Etc.

Quantitative data

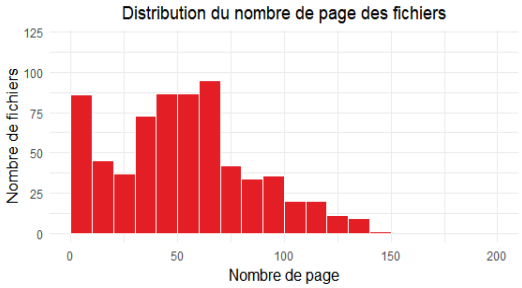
- Balance sheet;
- Own funds ;
- Technical provisions
- SCRs
- LAC, capital add-one;
- Other:
  - SCR sub modules;
  - Employees number;
  - Client number ;
  - Turnover.
- Etc.

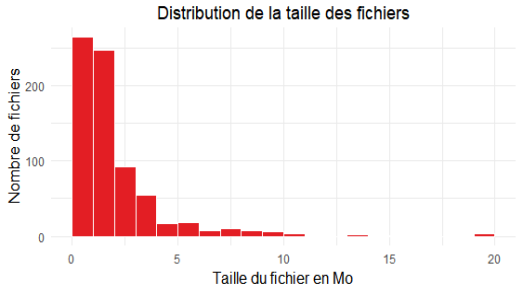
Qualitative data

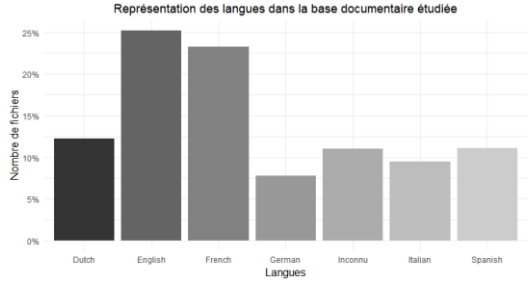
- LoB;
- Top management actions;
- Outsourcers, software, auditor;
- Currency;
- Accounting information;
- Reinsurers ;
- Reinsurance methods
- Etc.

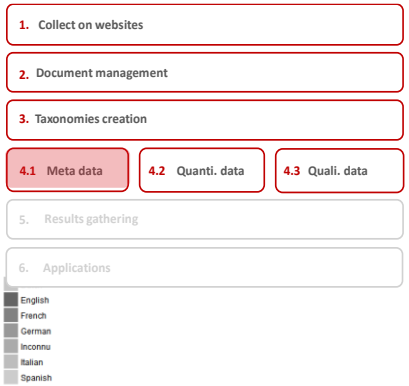
III. Results – Meta Data

- Use of meta data for different studies:

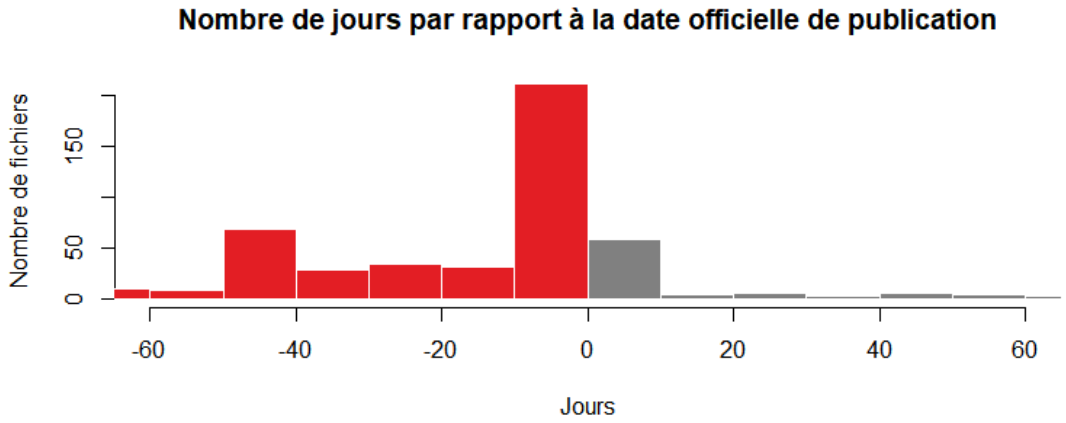


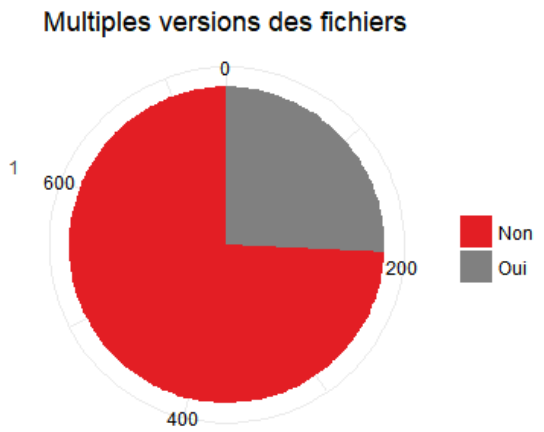






- « **Good and bad students** » observed through several charts (calendar and versioning):

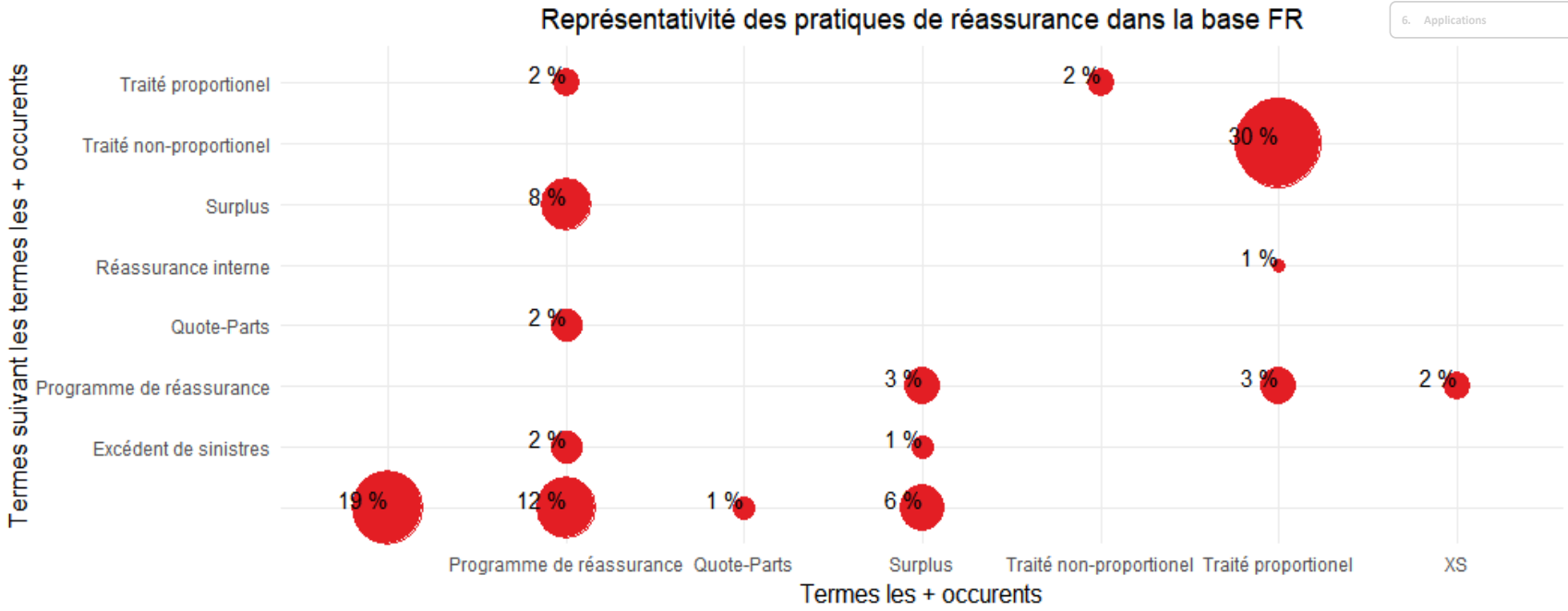




III. Results – Qualitative Data

1. Collect on websites
2. Document management
3. Taxonomies creation
4.1 Meta data | 4.2 Quanti. data | 4.3 Quali. data
5. Results gathering
6. Applications

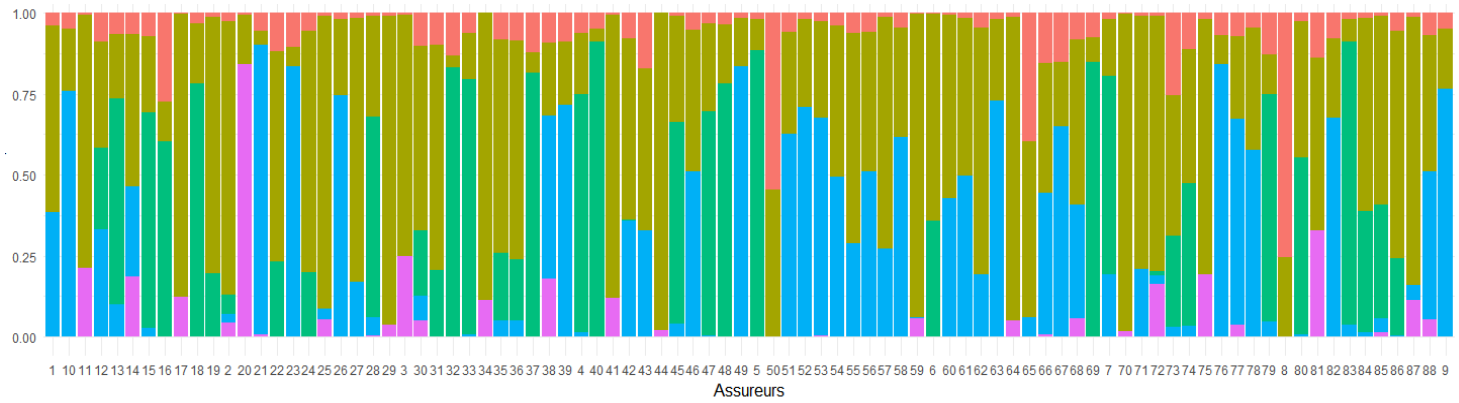
- A text analysis on the documentary data base allows to **understand market practices or market connections:**

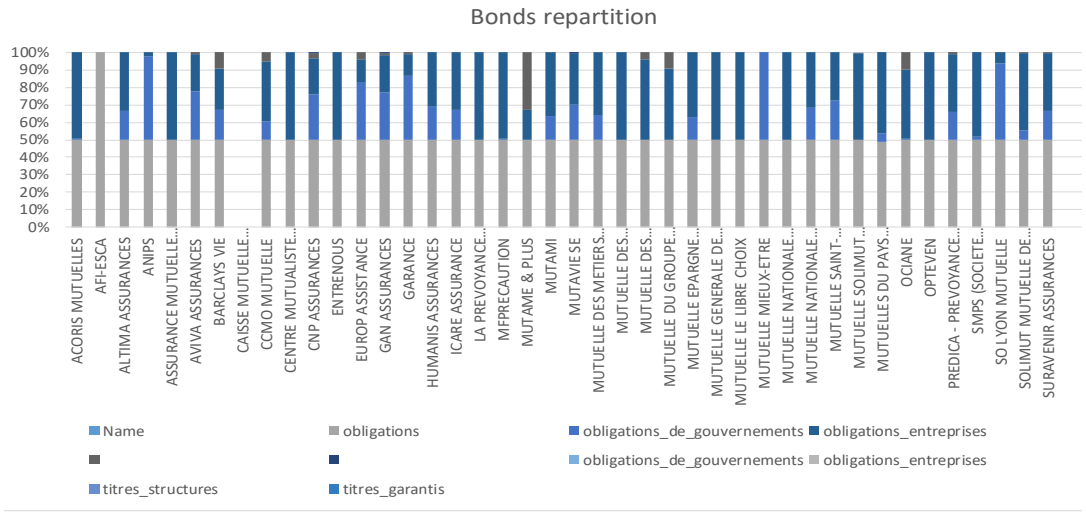


- And **many other possibilities**: modelling hypothesis, risk appetite approach, asset strategies, accounting methods, top management actions, etc.

III. Results – Quantitative data

- **Collect numerical values** such as SCR, MCR, number of clients, turnover, etc. but also of tables(QRTs like **S.02.01** and **S.25.01)**



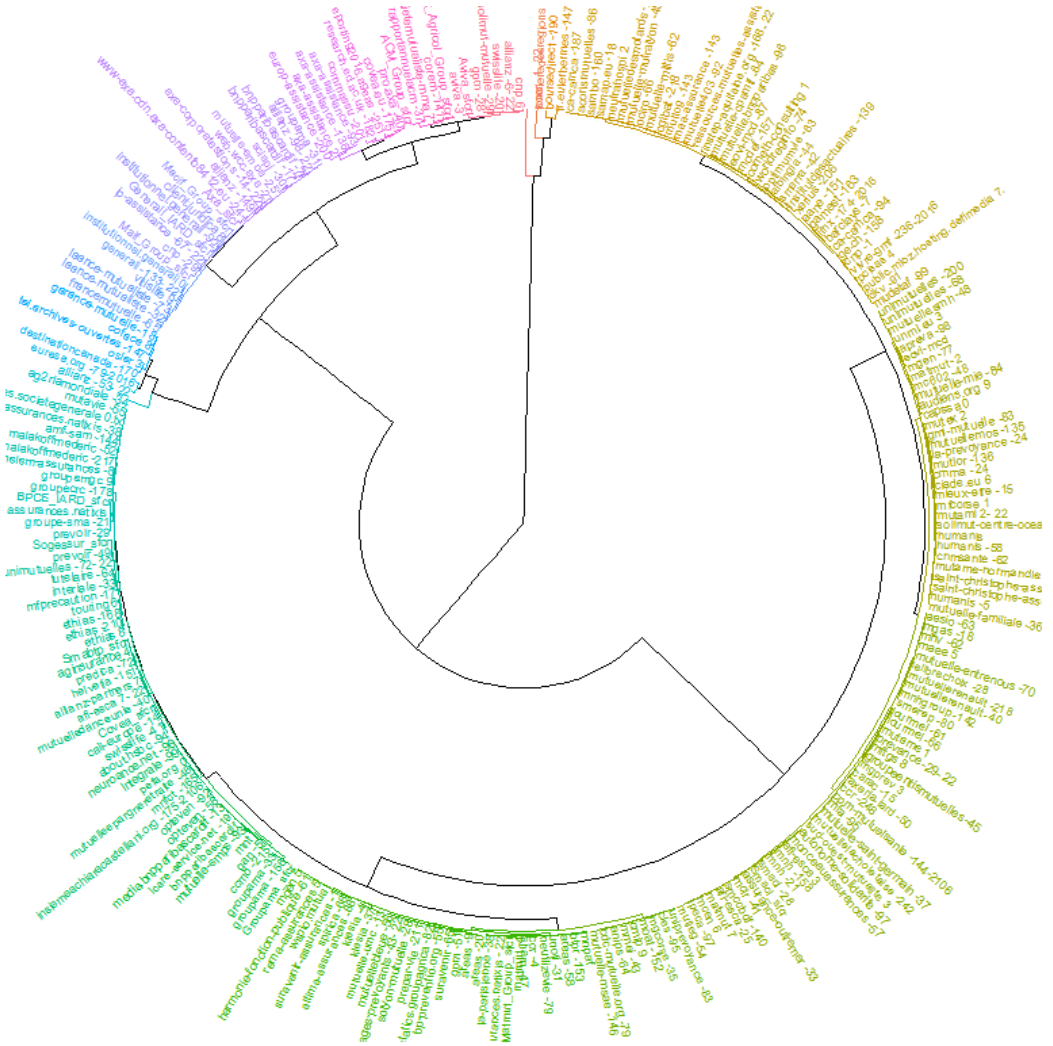


1. Collect on websites
2. Document management
3. Taxonomies creation
4.1 Meta data | 4.2 Quanti. data | 4.3 Quali. data
5. Results gathering
6. Applications

IV. Application – Case 1

- **Case 1**: Non supervised machine learning model to create custom clusters

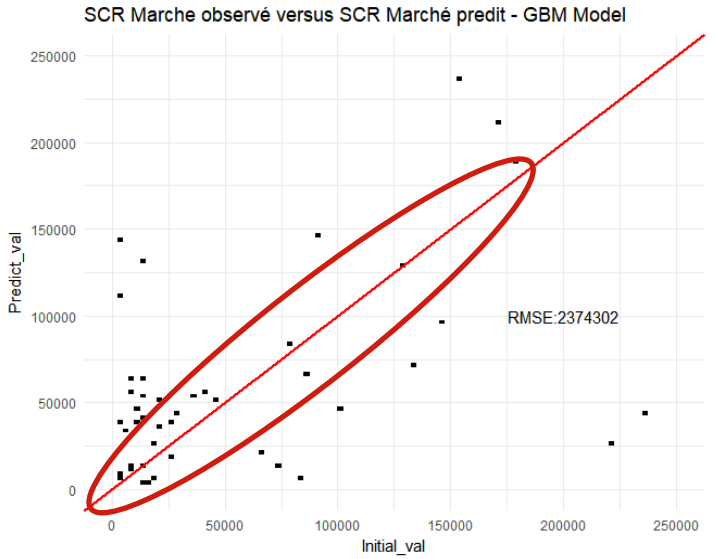
- Goals and motivation:
  - Allow to **make comparison** between companies **based on custom criteria**
  - For example: make cluster based on balance sheet and SCR and show discrimination due to reinsurance strategies

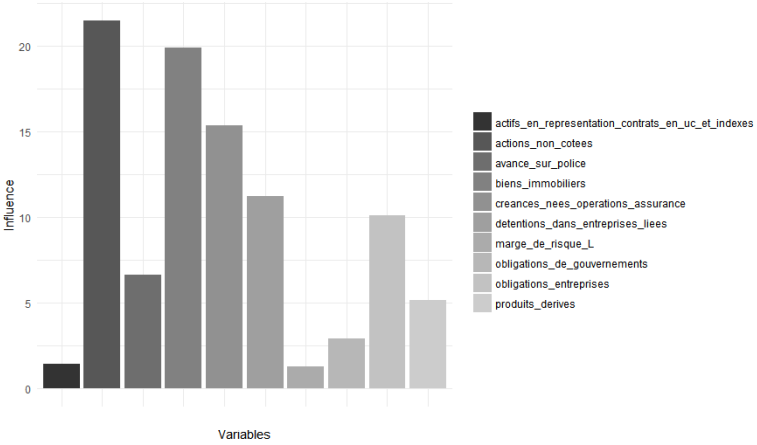


IV. Application – Case 2

- **Case 2**: supervised machine learning to predict estimated SCR values

- Goals and motivation:
  - Allow to **create a proxy to estimate SCR values** based on few variables
  - For example: propose to customers a tool that could help them to have more intuition regarding their SCR values and importance of drivers such as reinsurance





V. Limits

SFCR content (both quantitative and qualitative) is volatile

Scraping and parsing techniques are currently a subject of noisy debate

Performance on a simple machine is limited and requires discipline

Taxonomies creation requires time and material

- Other constraints:

- Currency and unit for amounts are not always easy to capture
- Companies set up protection not to find documents easily
- Database is still small to apply large ML models

V. Conclusion

- General remark: it is not **so easy to compare publications** even if the main characterictic of these documents is to be « comparable »

- 

Analyse and control easily — We have been able to create a process to **extract different types of information** (properties, labels, numbers) from **more than 1000 SFCR**

- 

Automate and save time — We have created an **automated and scalable process** using **open source material** that allows to produce structured datasets in **less than 6 hours**

- 

Create value for applications — We have developed **different business cases** that demonstrate all the potential of collecting unstructured data

Contact

Thank you !

Do you have questions ?

Aurélien Couloumy,
Head of Data Science
M FR +33 6 26 13 09 97
M BE +32 486 81 62 32
Aurelien.Couloumy@reacfin.com

Follow us on Linkedin:

Reacfin
Know-How to Risk

Reacfin Tooling
Reacfin Consulting
Reacfin Academy

About us

Reacfin is a consulting firm focused on setting up top quality, tailor-made risk management frameworks and offering state-of-art actuarial and financial techniques, methodologies and risk strategies.

Reacfin SA - Place de l'Université, 25 B-1348 Louvain-la-Neuve (Belgium) - 0032 0 10 84 07 50 - www.reacfin.com

Appendix – IT framework

- IT Language:

- Packages/Libraries (not exhaustive):

- tm, tidytext, udpipe, stringr, quanteda, etc.
- rvest, rcrawler, etc.
- caret, gbm, randomforest, factominer, etc.

- nltk, word2vec, doc2vec, etc.
- tensorflow, scikit learn, etc.
- pytesseract, opencv, etc.

- Database & Data visualization:

Local files (CSV/txt files) and remote ones (AWS)