*Date :* 17/03/2017

*Referring to :* Joint Committee Discussion Paper on the Use of Big Data by Financial Institutions

*Company :* CNP Assurances

*Address :* 4 Place Raoul Dautry, 75716 Paris Cedex 15, France

* *Question 1 page 12: “Do you agree with the above description of the Big Data phenomenon? If not, please explain why. Please also mention whether you consider that other characteristics are relevant to understanding the use of Big Data.”*

Answer:

The description of the Big Data phenomenon seems very complete. However, we have some remarks on the chapter "Definition and scope»:

* The importance of the two components: "data" and "model" seems to be clearly identified in both paragraphs 1 and 2. However, we propose to qualify the importance of 3 "Vs" in paragraph 1. Indeed, this 3 criteria are not mutually exclusive. Let us take the example of the first "V" concerning the volumetry of data. It is the most emblematic "V" among the 3 and yet its interpretation differs according to the type of industry concerned. Thus, for financial institutions, data volumes are considered important from several terabytes of data. In the case of insurance, this limit is commonly set at several hundred gigabytes of data. It is possible to obtain excellent results by combining the other 2 "V": Variety and Velocity with the models defined in paragraph 2 (especially machine learning models) even with only a few tens of gigabytes of data. Some actors will consider themselves to be in a Big Data context since they use varied data with Machine Learning algorithms whereas for others this will not be the case due to lack of volumetry. However, we are quite within the scope of this consultation document.

Therefore, in order to remain as broad as possible and allow us to represent as widely as possible all the actors involved in this consultation, we propose to introduce a sentence at the end of paragraph 1 in order to qualify the 3 criteria by specifying that they are not mutually exclusive in a Big Data environment.

* In Chapter 2, the term “correlation” is used twice to indicate that Machine Learning models are capable of detecting them. However, the notion of correlation is mathematically clearly defined and very easy to calculate: no need complex models to calculate it. The term correlation thus seems inappropriate in this context and we propose to delete it the two times it appears in the text. It should be noted that the term "patterns" that has always been associated with them perfectly describes the objective sought.
* *Question 3 page 13 : “Do you offer/are you considering using Big Data tools as part of your business model? If so, please briefly describe: i) what type of entity you are, e.g., long established, start-up, a product provider, an intermediary; ii) the service you provide; iii) the nature of your clients; iv) your business model; v) whether the Big Data tools/strategy were developed by an external company or internally and whether you have related agreements with other entities (including non-financial entities)23; vi) what are the types of data used (personal, anonymised, user data, statistical data etc.) sources of data; and vii) the size of your Big Data related activity and/or forecast activity (e.g. to what extent are business decisions already taken on the basis of Big Data analysis; what other business actions could be based on Big Data in the future)?”*

Answer:

CNP Assurances Group is the first life insurer in France, in particular for savings and loan insurance. It is also present internationally and in particular it is the 5th insurer in Brazil. CNP Assurances began its digital and Big Data transformation few years ago. CNP Assurances' Big Data strategy has been implemented in several orientations, both external (investment in start-ups, creation of a 100% digital insurance subsidiary) and internal (establishment of a department dedicated to the digital transformation, creation of a Data’Lab: pole of expertise in Data Science). Today, the implementation of Big Data and related techniques is no longer a future challenge for CNP Assurances but a reality : marketing targeting campaigns are based on the combined use of internal data and Open Data with Machine Learning algorithms and many management processes have been optimized using Big Data tools and Predictive models.

* *Question 7 page 20: “Do you consider any of these regulatory requirements as unjustified barriers preventing you from using Big Data technologies? If so, please explain why. Please also explain whether you consider that further regulation (including soft law/guidance, etc. and insofar as it falls within the scope/remit of the ESAs) should be introduced to facilitate the use of Big Data technologies.”*

Answer:

In general, we draw ESAs' attention to our concern that the gap between the EU's regulatory requirements and those of other less restrictive countries (in particular for the US) may create a competitive disadvantage for European companies in the global context in which we operate.

In the more specific framework of this "consultation paper", we would like to comment on the following passages:

*- "Consumers also have the right to know the logic involved in the automatic processing of data concerning them, in the case of the automated decisions" (paragraph 19)*

*- "data subject shall be informed about the existence of automated decision making, including profiling, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing" (paragraph 21)*

The regulations regarding these passages seem to us both too restrictive and potentially unsuitable for the Big Data. Thus, understanding the underlying logic of predictive algorithms to automate decisions is difficult to achieve. Indeed, these models are constructed to optimize decision-making and are very difficult to interpret: for a given data subject, the result of the models, even if it is deterministic, is obtained by matching hundreds (or thousands) indicators and / or variables that have undergone several transformations during the process, which does not allow a full understanding of the process that led to the decision. Thus, the requirement for information on the underlying logic of the models used for decision-making seems to us unsuitable for Big Data. It also seems too restrictive because strict and literal application of this requirement would provide to the data subject the thousands of elements that have been used in the decision-making process concerning him. But these thousands of elements are so many indications that, in some cases, could permit to reconstitute the model as a whole and thus to break the industrial secret. We are therefore alerting the regulator to the potential risk of a competitive disadvantage that could result from a strict application of this regulation as well as its inadequate aspect within the framework of the Big Data.

* *Question 19 page 36: “What are key success factors for a Big Data strategy (i.e. the adaptation of the business model/plan towards Big data driven technologies and methods)?”*

Answer:

A success factor for a Big Data strategy is, in our view, first and foremost a very personalized approach to everything that is being done in this field. Indeed, the data valorization within the framework of the Big Data is done in an entirely different way from one company to another, even for the same sector of activity and for the same problem because the approach to adopt must be adapted to the characteristics of each one. The exploitation of very diverse data is always done for searching optimization and with logic of performance which requires developing solutions each time different because they are optimized by integrating the specificities of each. In our view, Big Data does not allow for generic solutions and tools. From this point of view, another key element for the success of a Big Data strategy is to encourage and diversify internal and external initiatives in order to create solutions adapted to the company. Indeed, the absence of a generic solution necessarily implies a part of failure in the initiatives carried out and that is why the fact of multiplying them and diversifying them seems to be a good solution. It is in this context that the adoption of agile methods and development approaches by PoC (proof of concept: short and / or incomplete experimental realization to demonstrate the feasibility of an approach) that we also recommend have been developed in recent years. Even in the event of failure, the multiplication of initiatives and projects generally favors the development of the digital culture in the company and therefore improves its ability to adapt to this new environment.

* *Question 20 page 36: “What are the greatest future challenges in the development and implementation of Big Data strategies?”*

Answer:

In line with our previous answer, we believe that the major challenges companies will face in the years to come in implementing their Big Data strategy are different according to the size of the company. So the stakes for small companies and start-ups will have a lot to do with their ability to remain competitive in terms of technics and technology, whereas the challenges for the bigger companies will concern rather their ability to create cross-functionality between the teams in order to bring together the different expertise and also their ability to change their working methods and/or their organizations.