Cambridge, 20 February 2023

Office of the Corporate Secretary
FINRA 1735 K Street
NW Washington
DC 20006-1506

Attn of Ms Jennifer Piorko Mitchell

Subject: ***FINRA Machine-Readable Rulebook Initiative***

From: Regulatory Genome Project (RGP) at Cambridge University.

The Regulatory Genome Project is an initiative led by Cambridge Judge Business School at the University of Cambridge. Its mission is to develop and support the adoption of the Cambridge Regulatory Genome (CRG), an A.I. built and tested, machine-readable open-access taxonomy structure that enables regulatory obligations to be organised and compared across jurisdictions.

The RGP welcomes and praises FINRA's FIRST initiative and the knowledge and experience we have gained in developing the CRG means that we are particularly well placed to provide feedback on the key future questions raised by FINRA.

Other Regulators and Partnerships

Other Regulators are looking into developing similar solutions to FINRA's FIRST. Several cases have surfaced in the last few years, starting with the UK's FCA discussion on the topic dated 2017. Their commonality is the aim to use Machines to read and identify rules within a Regulator's policies body. However, FINRA should be praised for having published FIRST (first!) and having provided a real-life business case that both regulators and academic communities can look at with the highest interest.

Research on the topic suggests that open-source and open-access solutions are necessary to optimise interoperability. Whilst a Regulator may choose to provide an API, the fundamental need is for the supervised industry to access the taxonomies and use these common schemes. It can then build on them, with immediate benefits for automated compliance solutions beyond the activities linked to the reporting framework.

To strengthen FINRA's efforts, we suggest workings toward a full taxonomy in collaboration with other regulators, Industry and Academia. Our research indicates that regulatory taxonomy structures may differ for different experts elaborating on the same rules. Even in the context of NLP solutions building taxonomies, reconciliatory activities are required to reach a balance between obligatory requirements and the public understanding as well as usability by Machines.

Other Technology Solutions

The Cambridge Regulatory Genome (CRG) is a set of open-access, core taxonomies currently being developed at Cambridge University by the Regulatory Genome Project team.

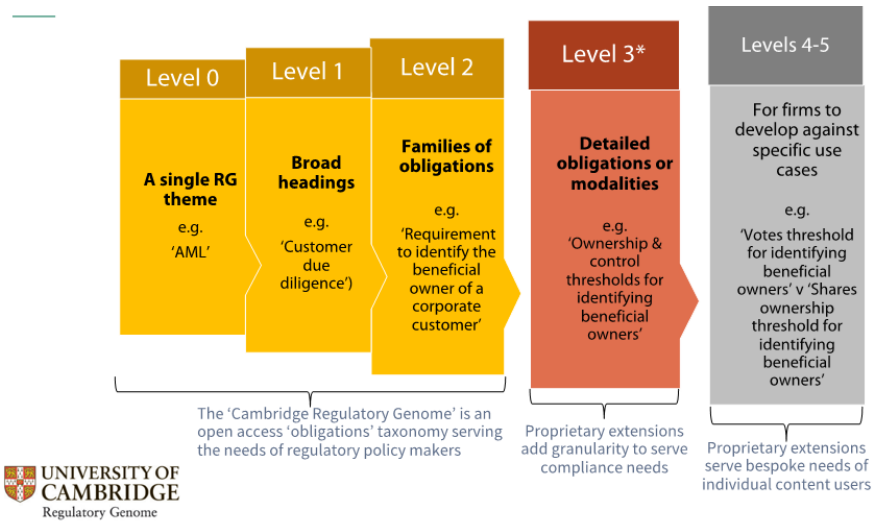
Whilst the CRG has been created to enable comparative analysis of financial regulations across jurisdictions, its methodology is based on the principles and guidance set by international Standards Setting Bodies. Hence, CRG taxonomies can be used by Regulators to embed international practice while developing machine readable solutions. They offer a base upon which Regulators can build a level three national-level taxonomy.

We invite FINRA to extend its current taxonomy with an 'overlayer' of the CRG, enabling international comparison, and consider connecting to the University to complete the set of tagging using A.I. based, NLP tested classifiers. Whilst initial testing is necessary, its precision would increase and avoid a more cumbersome manual approach over time. It will also enable a more productive environment with less biases from those in industry operating A.I. US based Industry participants would be welcome to collaborate in the process as FINRA sees fit.

Taxonomies Risks and Solutions

There are potential risks associated with multiple regulatory taxonomies being developed. These risks would equally affect Industry and Regulators, both at national and international level. On the first case, the Regulator would have increased disconnected communication and reporting frameworks with Firms, with increased internal costs and requirement to reconcile the information provided. On the latter, international cooperation and firms crossing borders would encounter the usual interpretative challenges.

Our analysis suggests that an approach of multilayer taxonomies can mitigate the proliferation of risks associated with multiple taxonomies, especially if Industry is looking to adopt A.I. solutions to connect to its API or leveraging FIRST's information structure.



Taxonomies built with a top-down family of structures reduce the risk of misinterpretation to the lower levels of the taxonomies while providing a common structure for the Regulator to operate on.

We strongly advise making the taxonomy available through an open-access manner for public-private collaboration and for further development of a FINRA taxonomy. In fact, our research suggests that taxonomies are key for harmonisation between regulators, firms and RegTechs.

The RGP team believes that FINRA should avoid the development of taxonomies within a silo or for internal purposes only. As for any other regulator, FINRA promotes transparency and clarification about its interpretation of standards. A tool such as FIRST is an extra tool for Industry and it is conducive toward a better adoption of standards and fulfilment of regulatory obligations.

Conclusions

We would encourage FINRA to dedicate a project to complete the taxonomy using machine learning techniques, in a format that provides sufficient granularity for all business cases while not detailing each single compliance case. Industry can contribute, for its own benefit, in the validation process of such taxonomies and validate the case for a FINRA full machine-readable rulebook.

Your sincerely:

Dr Giovanni Bandi
Executive Director, Regulatory Genome Project
Cambridge Judge Business School
University of Cambridge