

Advancing the SDGs: Understanding and Improving the Role of Corporate Transparency through Big Data

High-Level Roundtable Discussion,
July 16-17, 2018, São Paulo, SP

Meeting takeaways

Supported by the [Emerging Markets Sustainability Dialogues](#) (EMSD) program and the [Swiss Agency for Development and Cooperation](#) (SDC), the [Columbia Center on Sustainable Investment](#) (CCSI), [Sao Paulo Stock Exchange](#) (B3), and the [Center for Sustainability Studies of Funcacao Getulio Vargas](#) (FGVces) are partnering to understand how Big Data analytics and artificial intelligence can contribute to assessing how emerging market multinational enterprises (EMMNEs) impact the sustainable development goals (SDGs), and how different stakeholders can drive corporate performance in this area.

The **second roundtable** (RT2) as part of this initiative brought different stakeholders together to discuss **three key questions**:

- How are Big Data and Artificial Intelligence (BDAI) technologies being used to better understand firms' policies and impacts on the SDGs and 2030 Agenda?
- How could BDAI technologies be used to expand and improve the collection and analysis of ESG information?
- How can companies, investors, and other stakeholders leverage the power of BDAI technologies to overcome their need for better data and make sound decisions to shape or respond to the sustainability performance of firms?

The discussions in this second roundtable focused, although not exclusively, on emerging market firms and, in particular, on the Brazilian context,¹ the ISE² and the information demands of its various stakeholders.

In this document, “Big Data” refers to data and associated processes whose scale, distribution, diversity and/or creation speed requires the use of new storage and analysis technologies to capture their value (adapted from EMC, 2013). Technological advances have allowed for a massive generation of data, from sources such as unstructured text, transactions and sensors. The term *data* refers to the “raw material” collected, which, once processed to gain meaning, will be referred to as *information*.

Core conclusions from the discussions were:

¹ In this context Brazilian multinationals were used as a case-study for other emerging market companies.

² The corporate sustainability index of B3, the Sao Paulo stock exchange: www.isebvmf.com.br



On the context of sustainable investments:

- There seems to be a lack of useful ESG/sustainability-related information on emerging market firms. Fulfilling investors', governments' and civil society's information needs is overwhelming to most companies; moreover, there are disincentives for firms to provide information that is negative. And even when companies are open to provide such information, making sense of it a challenge to most stakeholders.
- The lack of quality data presents a paradox: How can there be such an unmet need given the large volume of data? A central part of the problem is that 80% of the newly created data is unstructured and cannot be processed with traditional data analysis tools. Big data is revolutionizing the way data can be gathered (in large volumes) processed and analyzed.
- However, even with such potential, the lack of standardization in reporting makes it difficult for machines to fully read and extract information from company reports. Reports that are meant to be visually appealing and use diagrams and other visual resources to convey information to human readers make reports difficult, or almost impossible, for a machine to read. A simple solution for this particular problem could be adding the same information structured as tables at the end or as appendixes to companies' reports. Although tables are not easy for humans to read they are preferable for machines.
- When it comes to Big Data, there are standard processes data scientists go through: 1) scraping sources and gathering data, 2) cleaning, structuring, and extracting relevant data, 3) analyzing the data and producing insights, 4) curating insights, often by including human analysis or supervision on top of machine produced analysis, and 5) making sense of and sharing insights. Although similar, there are different approaches when developing algorithms for each step; and those different algorithms can create very different outputs. Some companies rely heavily on advanced quantitative models to select relevant data while others prefer using analysts to review the information and decide what is important, instead of a machine.
- Big Data and associated analytical technologies also allows a different timing: in comparison to traditional data providers that are looking at largely annual data, many are now looking into new processes that are more timely and less reliant upon annual reviews. Since Big Data allows for continuous real-time monitoring, it may also be applied by users to anticipate emerging trends or issues before they are officially reported by companies in annual filings. As not all stakeholders have the same need for timely updates, the relevance of the data time lag is contingent upon the goal (what will they do with the data) of the stakeholder.
- Different stakeholders have different abilities to access data. Processes for collecting and analyzing data are labor-intensive and expensive).

On the lack of social and environmental data:

- As discussed in the first roundtable (RT1), environmental and social data are the hardest to collect and use in assessing companies' ESG performance. Important reasons for that are the diverse nature of such information and the lack of standards for more structured collection, management, disclosure and assessment of existing data. Many believe that much of the information needed for assessment already exists, but is in an unstructured format, and dispersed in diverse locations. For example, it is likely that local level newspapers or bloggers collect and report environmental and social data because relevant information is easier to spot and very relevant at the local level. Such information may not be reported in national mainstream media.
- As some ESG information providing/assessment companies are seeking to find better ways to capture and measure the environmental and social aspects of companies' performance, they may partner with local civil society organizations. Nevertheless, there remains the challenge of how to process and quantify unstructured social and environmental information, as well as how to assess its relevance and reliability.

On data about parent and affiliated companies:

- At this point, even with the use of Big Data, it remains difficult to connect data from parent and affiliated companies, especially when the economic group is complex and/or not clearly communicated. According to one participant, "It is impracticable to keep track of large companies with hundreds of affiliates." Nevertheless, if one cannot link the activities of affiliates with activities of the corporate parent, and firms are able distance themselves from their relatively poor-performing affiliates, those firms are able to "greenwash" their operations and provide misleading pictures of their policies, practices and impacts.

On the intersection between the SDGs, ESG, and Big Data:

- A key discussion from RT1 was about the gaps between existing ESG-oriented tools and frameworks, on the one hand, and the SDGs and Agenda 2030, on the other. In particular, the information relevant to assessing firms' contributions to the SDGs may be broader than the information material to investors; one may want information from companies even if it is not (or not yet) necessarily deemed legally material to investors' decisions.
- Many existing ESG reporting and assessment tools and frameworks are evaluating how their approaches compare/align with the SDGs, and mapping their frameworks against the SDGs. One issue that arises from using existing frameworks is that they do not fully capture the essence of the SDGs, for

example, when SASB applied its materiality map to the SDGs, does that mean that only that which is material is reported thus leading companies to report on some but not all SDGs?

- Some participants of RT2 argue that we should have new indicators derived from Big Data and compare them with traditional indicators, in order to better understand the information that is being processed.
- A reason for this apparently conflicting perspectives is that Big Data, by allowing processes that would be not feasible in other circumstances, opens completely new avenues for investigation and, therefore, for the creation of very different indicators.
- A possible corollary of this debate is that the way forward of corporate sustainability assessment should include (i) improvement of existing ESG tools in order to fully incorporate the SDG perspective, using Big Data as a means to enrich their information basis while alleviating the reporting burden for companies; and (ii) the creation of innovative approaches to understand and assess companies' sustainability, harnessed on "Big Data thinking" and conceived under an integrated ESG/SDG framework.
- These approaches seems to be complementary, rather than antagonistic, although there may be concerns regarding possible competition for the attention of companies and their stakeholders.

On the issue of capacity building

- To seize the full potential of Big Data, employees from MNE's and financial analysts related to sustainability management, assessment and reporting should be trained in order to learn how to think and work with new technologies efficiently and ethically.
- Technology professionals should similarly receive training in order to understand the connections between their field of knowledge and sustainability, and advance efforts to identify innovative routes and opportunities.

On reliability/credibility of information sourced through Big Data:

- Big data does not provide "perfect or neutral information", because it depends on assumptions provided by humans to the machines. The quality of the information produced depends on how people teach bots to search and to make sense of data.
- Also, the quality of the information provided for machine analysis have direct influence on its outputs: the analysis based on non-reliable sources will be as non-reliable as the sources.



- Therefore, careful human planning, supervision and curatorship of inputs, analytical frameworks and outputs are still essential, even with the use of advanced technology.

Which data providing company is the best?

- Although very basic, this question cannot be answered without a clear understanding of what kind of data a user wants and how it will consume it. For example, some users are more interested in daily or even real-time data updates, while others are more concerned with the comprehensiveness and depth of data periodically available.
- Additionally, users approach data providers with different perspectives and needs regarding sustainability issues, causing data providers to adopt very different philosophies in their solutions. 'Data providers' entire use of the data therefore may go through different processes because it is conducted from different perspectives; there is no best approach, only different perspectives. For example, some have a qualitative approach and look into the meaning of each report given a certain context while others use statistics to get the bigger picture and do not spend much time addressing the specificities of each data point. A comparative study will be produced especially on this topic.
- As noted in RT1, and once again confirmed by different stakeholders, investors and companies or professionals advising them, usually look for (i) a single and comparable score that summarizes a complex issue to decide between different investment alternatives, and/or (ii) a structured and well-grounded analytical framework that strongly supports the high-level indicators and justifies conclusions, recommendations and decisions.
- The complexity, diversity and broadness of sustainability matters make it very difficult to meet such demands, but Big Data can help by improving and enriching the processes of data gathering and analysis process.
- An important aspect is that most of these services are not available for free, meaning that (i) investors with resources can access the data, but the ability of other stakeholders to do so is limited; and (ii) the information geared towards investors interests may not be the information desired/needed by other stakeholders. Therefore, solutions should be identified to level the play field and avoid widening the information and power gap between societal sectors.

The language issue:

- Most existing artificial intelligence (AI) tools developed to work with sustainability issues have been developed in English. Since a core part of the sense-making process of Big Data depends on natural language interpretation, the language issue is central for the effectiveness of each solution.
- In the debates at RT2, different strategies adopted by information providing companies were identified: (i) processing only materials originally written in the AI



system's "native" language (usually English); (ii) broadening the universe of sources, by translating materials in other languages to the AI system's "native language" (usually, from other languages to English, using electronic translating systems); (iii) "teaching" a new language to an existing AI system, which means making use of the basic logics and algorithms, but creating from scratch the actual creation of lexica and dictionaries in the new/local language.

- There are various trade-offs between these paths, including their cost-effectiveness, the time required to have a usable output, the quality of the text interpretation, and the quality of the final outputs. The assessment of such trade-offs is very context-specific, depending largely on the similarity of the relevant languages, the availability of original sources in English, the particularities of the themes being analyzed (jargon and context-specific terminology and meaning), and on how critical precision in language interpretation is for the relevant user's needs and desired outputs.
- A critical aspect to be considered is the importance of accumulation and quantity of data available in BDAI applications: the bigger the base is, the better are the results. Therefore, the existing systems, originally in English, have an important handicap.

To move forward, it is important to:

- Identify one major player (e.g. sovereign fund) in Brazil that could acquire existing technology and create a demand for the use of BDAI. In other words, identify a market mover or leader that could kick-start this process of improving NLP in Portuguese. The Brazilian Development Bank (BNDES) and FEBRABAN (The Brazilian Federation of Banks) were identified as potential major players that should be consulted. The International Finance Corporation (IFC) could also play a role in this process, due not only to its core activities, but also due to its relationship with ISE, since IFC have provided the seed money for its creation, 14 years ago, and is also a member of ISE's board.
- Another important output is to have an investor's guide on overcoming the data scarcity in EM through Big Data. The guide would assist those seeking to make responsible investments by presenting the discussion on SDGs and ESGs, the quality and gaps of methodologies, and how investors can leverage their ESG/SDGs investments using BDAI Benchmark of Big Data Providers.
- Lastly, in the past two years both the Santiago and Buenos Aires stock exchange inaugurated their own corporate sustainability index. The Brazilian corporate sustainability index (ISE) was studied by both cities while designing their own index. Another output is to create a material for other stock exchanges sharing many key lessons and thoughtful questions that one must consider while designing a sustainability index. The material would be used to strengthen SDG contributions in EM.