

Fishing for Carbon Solutions in the Global Data Lake

Paul A. Davies of the law firm Latham & Watkins' London office, who moderated the Environmental Law Institute's 8th GreenTech webinar—"Using the World's Data Lake to Solve Global Challenges"—opened the event on an upbeat note. In the wake of COP26, the world's "vastly underutilized" data could be combined to create a more "predictive approach" to solving climate change and other global environmental problems, Davies said. He then introduced the webinar's two expert speakers: Mark Bernstein, the co-founder and President of nonprofit Earthshot that is inspiring people to act on climate change through video gaming, and Toby Sykes, Head of Data at Altruistiq, a company begun in 2021 to help companies measure and reduce their carbon footprint.

Addressing the question of what data users actually want to solve and whether the data lake can provide the tools, Sykes explained that "vast opportunities" exist for private sector decarbonization across the value chain if companies use the right tracking tools and measurements. Drawing on more granular emissions data and strong analytical capabilities, companies can set clearer targets and find numerous emissions-cutting opportunities. A significant challenge for organizations, however, is to achieve enough automation so that teams don't spend the bulk of their time gathering data.

For Bernstein, data must be used to convince companies to find climate change solutions, to help individuals identify their opportunities, and to assist investors in deciding where to invest their funds. Sykes noted that J.P. Morgan had recently announced it will invest \$2.5 trillion over 10 years on climate change and sustainability and will need help in deciding where to focus those dollars.

A breadth of data types are all part of the massive data lake that is now available to define practical solutions in ways that individuals and institutions can understand, including technical data used in studies and calculations and data found in traditional and social media. For example, people have taken numerous photographs of Glacier National Park and posted them on Instagram. Combining those images with traditional data can facilitate the recognition of trends, thereby enabling a more predictive understanding of problems and their solutions.

Although the data lake will be critical in solving biodiversity and other environmental problems besides climate change, biodiversity solutions are harder because proving causation in this area is more challenging, Sykes noted. For example, it would be harder to determine the impacts of a manufacturing plant on surrounding biodiversity. But it would be absolutely critical to track changes to the surrounding biodiversity over time, he said, noting that platforms have been deployed to record bird and other wildlife movements and satellite data are being used to track deforestation. Additionally, Bernstein emphasized the critical link between satellite imagery and data "on the ground" to understand, for example, what is occurring in coral reefs. In addition to climate change and biodiversity, such data linkages can help track and solve local air quality issues.

To help make Greentech that is associated with the data lake more commercially viable, standardization is crucial, especially in the area of carbon accounting, but standards should not be overly rigid, Sykes said. The long-standing Greenhouse Gas Protocol is a great starting place, but it presents guidance, not required standards. Moreover, businesses find it challenging to map their various activities to the Protocol's Scope 1,2, and 3 structure. At the COP26 meeting, the IFRS Foundation announced the formation of a new International Sustainability Standards Board to drive better carbon tracking and reporting, and that effort should produce greater consistency in the data generated by the private sector. Both speakers emphasized the data lake's exciting potential to help carbon management strategies move from a reactive to a more predictive mode. But, Sykes reiterated, inconsistent standards and metrics act as a barrier to predictive capabilities. Stronger guidance on what is permissible and on appropriate data uses would accelerate the emergence of predictive carbon management.

Bernstein raised the issue of data verification, stressing its importance to confirm the reality of suspected problems. Perhaps blockchain could be used someday for verification. But lack of investment is holding back the use of such digital tools. In addition, machine learning and AI are progressing, but a greater push on such technologies is needed to develop their capability for helping make decisions. Sykes commented that blockchain has a definite role to play in certifying carbon reductions and building an "immutable ledger" or "robust audit trail" of emission reductions achieved through initiatives, but in his own work, so far, he has not yet encountered blockchain. As a data ecosystem gets built across multiple organizations—and the complex supply chains that impact their carbon footprint—blockchain will be a "key enabler" to manage information flows in a fully transparent way that assures accountability for emissions reductions. Among the other technologies most beneficial for facilitating better data use are "open banking," the Internet of Things, and cloud processing.

Bernstein commented that the tools to find, sort, and verify the large amount of information include machine-learning searches and "scraping" patent and social media data, but "you do need people" for the necessary machine-people interaction. Digital tools can collect and organize data, but people are needed to verify the results. Some groups are working on improved tools for recognizing information, including, for example, a company whose employees are conducting machine-aided searches to find and recover ocean plastics.

Given concerns about the potential bias in algorithms as AI is increasingly used to "interrogate the data lake," Sykes emphasized the need to ensure full transparency regarding how an algorithm makes its choices. It can be challenging to get "into the black box," but extensive research is being conducted on how to ensure the right controls are adopted and AI is "a force for good" that can be used without adverse effects. Bernstein added that diverse people from around the world are needed to help sort, verify, and analyze data. To maximize the data lake's value for all users, Bernstein's organization is working to make its portal easily searchable and its information readily understandable by users in different locations, including companies with locations around the world whose businesses differ. Sykes also noted the challenges of transparently tracing the lineage of data and of ensuring the right data are used for specific calculations.

Regarding questions of data availability and limitations, Sykes noted that there is little data sharing between competing logistics companies, although some sharing occurs. But his company is working to generate more sharing. Altruistiq has found that organizations are willing to provide their data, but getting data from suppliers on behalf of corporate customers can be a longer process. Bernstein noted that it would be ideal if sustainability solutions were widely shared but that has not happened. A tension exists among businesses aiming to foster positive change but also trying to make a profit that stands in the way of everyone having ready access to all available information.

Specifically commenting on access to drone data, Sykes reiterated that depending on the answer being sought, less detailed data might be needed. Bernstein noted that drone data for analyzing land use, water availability, soil carbon on farms, and other such purposes is very helpful and is cheaper than Lidar flyovers. But companies that invest in using drones must be assured a return on their expenditures. Government support and policies—such as incentives and carefully targeted regulations—can help, but growing consumer demand for more information before they buy products will be more effective than government policy.

In remarks on COP26, Bernstein said that countries' methane control commitments will lead to investments in measurement tools to support independent verification. Sykes commented that perfect data are not essential, as long as the data are good enough to recognize the general yearly trend, whether emissions are going up or down. If asked to advise governments on better ways to use the data lake for meeting the COP26 challenges, Bernstein said understandable information to make better choices is essential. For Sykes, the key is to promote better investing in the data ecosystem and in agricultural tech and other climate-related innovation, along with greater standardization to make the data widely usable.

Bernstein mentioned Earthshot's goal of presenting a clean-tech future to the 2.5 billion gamers around the world, age 18-33 years old. He expressed enthusiasm for an idea suggested by an attendee of using games to help catch illegal loggers. Satellite imagery could be used, with gamers competing with each other to find illegal logging and alert authorities. Games can drive engagement and understanding, Sykes added.

In 2018, the Environmental Law Institute developed a game to educate the public on coastal resilience. Cards Against Calamity is free and can be played here:

<https://cards-against-calamity.org/>