Session Three:

Alternative Data

CU SEAS IEOR 4723:

Financial Eng. for ESG Finance
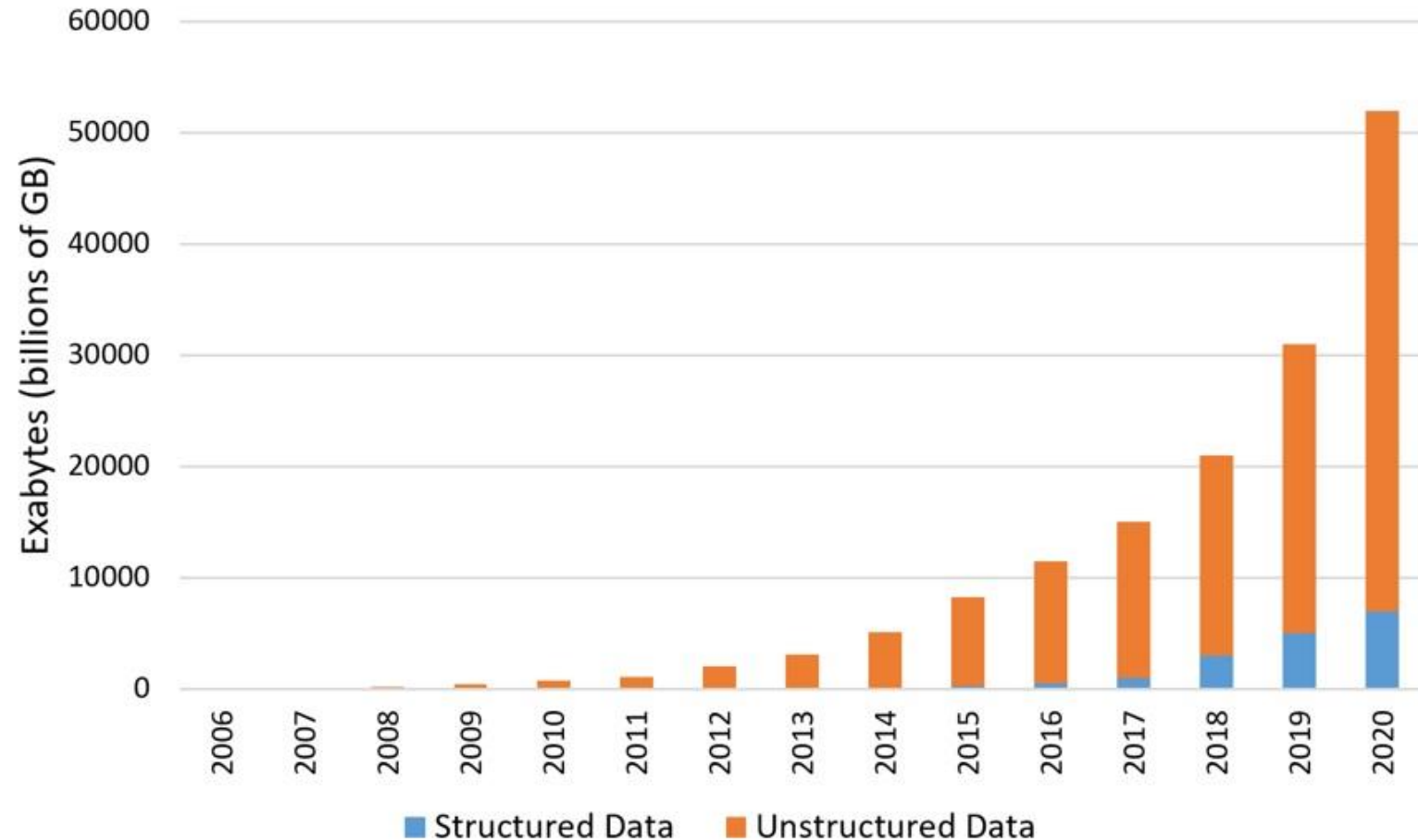
3.1 *Definition of Alternative Data and Their ESG Applications*

Data Drives Financial Decisions…

As the amount of overall "data" available to an investor grows rapidly, the proportional share of "unstructured data" also increases

ESG Finance has an even stronger reliance on unstructured or "alternative" data, because the traditional structured sources for ESG data are – at the moment – still very inconsistent. Due to the current lack of reporting standards, investors are left to rely on "alternative" data sources

The Cambrian Explosion...of Data

Y-axis: Exabytes (billions of GB) — 0, 10000, 20000, 30000, 40000, 50000, 60000

X-axis: 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020

Legend: ■ Structured Data ■ Unstructured Data

Source: TBD

Data Providers

Fintech startups have taken the lead in offering both ESG rating (discussed previously) and underlying data used in ratings' generation

A few prominent examples include (see also *Datarade AI*'s "top ten" lists discussed next, featuring the same names):

- *TruValue Labs,* now a FactSet Company, applies AI to quantify Environmental, Social, and Governance (ESG) data found in unstructured text sources including news, trade journals, NGO, and industry reports to deliver transparent, timely, consistent ESG data to investment professionals

- *OWL Analytics,* an alternative data company that focuses on environmental, social, and governance (ESG) research. Helps investors integrate ESG data into their investment solutions with the goal of delivering enhanced investment results and positive social impact compared to benchmarks

- *Arabseque,* a group of financial technology companies offering sustainable investment, advisory, and data services through advanced ESG and AI capabilities

Some of the new offerings were either acquired or developed in-house by more traditional data providers (Bloomberg, Reuters Eikon, etc.) as well

Alternative Data

Alternative data can be roughly categorized into four types:

- **Sentiment**: Social-media feeds, news flow, corporate announcements and other published media are monitored and analyzed for clues to sentiment on stocks, products, and the economy. This category includes the analysis of language used by executives on earnings calls, etc.

- **Other Web Scraping**: Involves compiling data from targeted websites in a bid to gain information on brands and products. Includes job listings and employee-satisfaction rankings, which can offer clues to a company's growth prospects

- **Credit Card Data**: Some data providers put together large panels of consumers who agree to share their credit- and debit-card activity. Panels made up of more than 3 million consumers are considered big enough to be useful; these data are used for real-time tracking of retail revenue

- **Satellites and Aerial Surveillance**: Used to track ships en route, monitor crops, and detect activity in ports and oil fields. Also, cars in parking lots are often counted as a proxy for retailer sales activity

Alternative Data Clearinghouse. *Datarade AI,* a commercial clearing house for trading in alternative data sets, hosts a range of ESG datasets. Useful data descriptions and categorization are laid out in their FAQ section (reviewed next)

Environmental Features in ESG Datasets

The environmental features in ESG databases are points that measure a company's impact on the environment. This includes waste disposal, sustainability of resources, and energy sources. Environmental features measure how the practices of a particular company impact the environment, and how a company intends to preserve the natural world. Since the 2020 Paris Agreement there has been an emphasis on how corporates contribute to cutting GHG emissions. Accordingly, the importance of a company's ESG environmental score has increased. Environmental features consider the sustainability of resources and energy sources, as well as contribution to pollution/waste. Key sample questions include:

Does a company use sustainable resources and energy sources?	• The amount of greenhouse gases, like carbon dioxide and methane, that a company produces can be tracked and recorded. The worse the greenhouse pollution, the lower the ESG score for this attribute. However, if companies move away from the use of fossil fuels, adopt a more eco-friendly approach and increase the percentage of power generated through the use of renewable energy, a company's ESG score can improve
Does a company produce greenhouse gases?	• The amount of greenhouse gases, like carbon dioxide and methane, that a company produces can be tracked and recorded. The more greenhouse pollution created by a company, the lower its ESG score. Companies that are heavily invested in exploiting carbon-based resources like coal, oil, and gas tend to be penalized in ESG scoring as it is difficult to sustainably acquire and process these resources
How much water does a company use and how is it sourced?	• The amount of water used and polluted by a company can also affect its ESG score. Nestle's ethics came in question when they were found practicing unethical ways of affecting water sources of African villages to peddle a particular baby formula that was discouraging breastfeeding
Does the company use renewable energy?	• Positive steps taken by companies to become more eco friendly, like purchasing a percentage of their power from renewable energy sources, can have a positive impact on the company's ESG score

Social Features in ESG Datasets

Social features in ESG data give insights into issues surrounding human capital, labor standards, privacy, data security, and stakeholder opposition of the company. The social features in ESG data consider business relationships of a company, providing insight into data concerning company employees, customers and the wider communities a company is based in

Human Capital Diversity

- Diversity within a company impacts its ESG data score and recent studies have suggested that there is a strong connection between the level of diversity and financial prosperity within a company. Wider diversity within a company can help drive innovation and is consequently considered a primary indicator of a company's overall health and ESG data score. Having a non-discriminatory and racially diverse organization is often a goal of a company. As companies become more global, ensuring work opportunities are available for a diverse range of individuals is becoming increasingly important. Equally, it is no secret that more companies are seeking to achieve a gender balance across their organizations. The ratio of male to female, and the inclusion of non-binary genders, has been an important factor in many offices' decision-making processes and companies often strive for equitable hiring practices to manage this. However, across a number of sectors, there is a staggering lack of diversity across its employees. This is especially the case when considering those in high level positions. A study found that globally, in 2020 only 9% of CEOs were female. This points to the fact that across many companies, levels of diversity may not be up to standard. This lack of diversity negatively impacts a company's ESG score, data which can be used to inform investment decisions

Human Rights

- As well as company diversity, the social attributes of a company's ESG dataset is determined by an organization's consideration for human rights. One way a company's social attributes can be assessed is through the consideration of an organization's Employee Turnover Rate. How long employees stay with a company and their overall satisfaction rate can impact a company's ESG score. These company datasets may offer an insight into how the company treats its employees. However, the human rights element of a company doesn't only consider the company's internal business strategies and relations but also assesses how a company treats the communities in which it is based and whether a company positively impacts the wider society

Governance Features in ESG Datasets

Governance features in ESG data concern a company's actions surrounding the distribution of the decision-making processes, taking into account corruption, business ethics, a company's board and management structure, and its employee relationships including employee payments and compensation strategies

Structure of Management

- The equity valuation of a company looks at the internal set of protocols of how the management across a company is structured. The Structure of Management considers the balance of power between a company's CEOs and its Board of Directors

Improving Employee Relationships

- In establishing a fair recruiting process that incorporates diversity, establishing values and policies that create a safe workplace is crucial. The Fortune 100 best companies to work for has become a coveted list for employers as this drives the hiring of valuable talent and has an impact on the growth of the company as a result. This also considers the ability of workers to have representation or a union

Compensation of Top Executives

- In recent times the remuneration and bonus packages of board members and top management employees have become a point of scrutiny by shareholders as well as equity investors

Fair Compensation of Employees

- Employers are liable to pay the employees a considerable amount pertaining to their designation. All major companies practice minimum wage standards. Actively working to close the wage disparity between genders is another important benchmark. Pay equity audits are available for auditing and they are in many cases made available to the public forums, which open up the ethics of the company to be questioned by investors

What Are The Sources of ESG Data?

ESG data comes from a variety of sources that include annual company reports, government reports, news reports, and social media posts

The majority of raw ESG data come from self-reporting sources; however, data can also be sourced from third parties and real-time signals
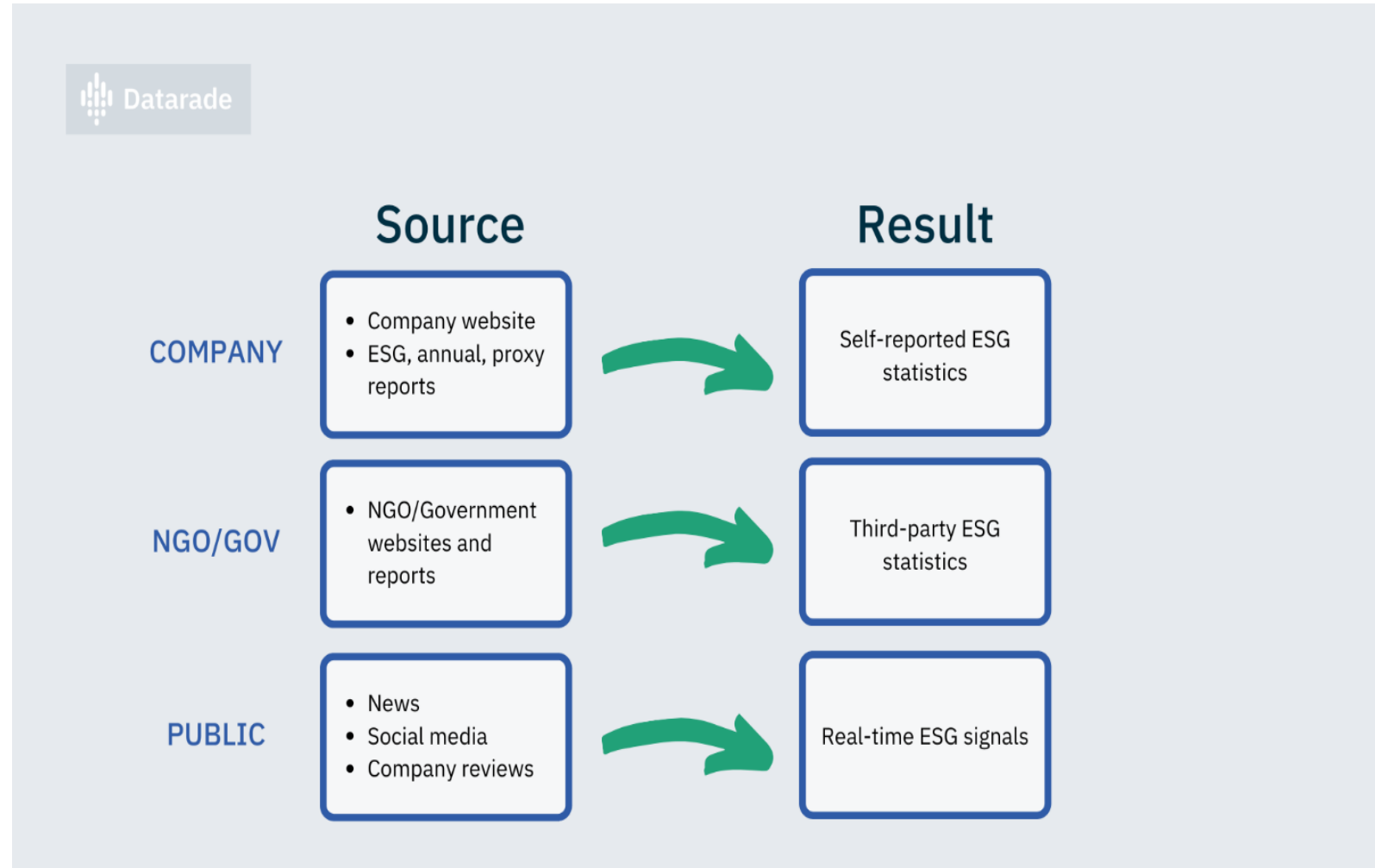
Source: Datarade AI

How to Find The Best ESG Data Providers?

Investors stress comparability and consistency in the data that is available in the market to construct an effective dataset based on which to score companies

As mentioned, a combination of a lack of standardization and a large amount of red tape makes identifying top providers difficult.

The providers' job isn't any easier. ESG sustainability reporting is not yet mandatory in all countries. Therefore, self-reporting companies are free to decide which ESG factors are reported; selective data are being disclosed to the public, making the data biased.

Lack of consistent and reliable information is a challenge that governments, private investors, companies, and the public have to face. ESG data providers are crucial in bridging the gap: they collect and evaluate information related to a company's ESG practice, and rate companies based on relevant factors.

Wider availability of data has brought increased interest in ESG investing, as data allow investors to conduct their own ESG research. Fortunately, hundreds of vendors are available, many of them ESG exclusive…

"Top 10 Data Providers" Lists

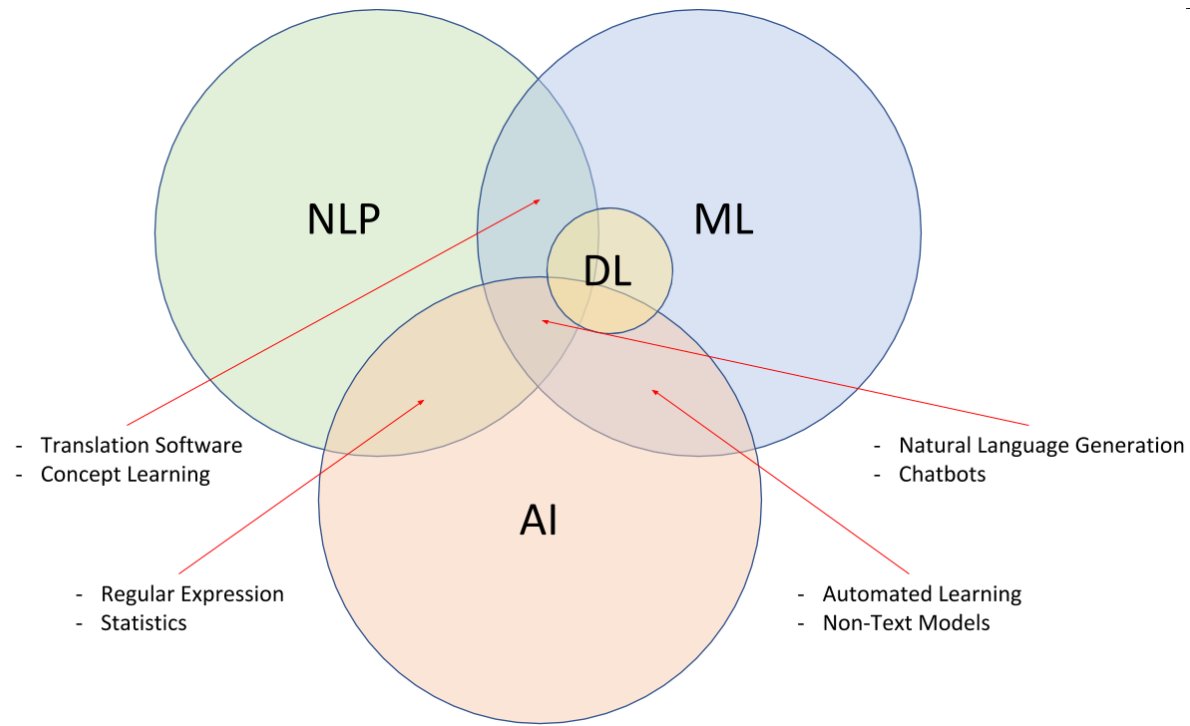*Below is the "10 Best Data Providers for ESG Investing" list from Datarade (not independently verified)*

- *TruValue Labs* provide data-driven solutions to enhance investment decision-making. They collect data from millions of sources per month, so that investors can manage risk and monitor portfolios across a range of countries and industries
- *Owl Analytics* is an ESG data provider. Intelligence supplied by Owl Analytics is causing the corporate investment landscape to become more value-based and sustainable. Owl Analytics' data allows investors to carry out ethical trading and drive up alpha generation
- *Arabesque S-Ray* was born out of the company's own need for powerful ESG metrics. Their ESG data assigns over 7000 companies with ESG scores based on 30,000 news sources and 700 NGO campaigns
- *Accern* utilizes AI and ML to power data-driven sustainable investment. The company offers historical ESG data covering the past 15 years, collected from Dow Jones News Wires and EDGAR SEC Filings
- *Sustainalytics* supports sustainable investment and enables traders to benchmark portfolios from over 4000 companies, based on detailed ESG metrics
- *Clarity AI* is a data provider that supplies intelligence for sustainable investing. The company leverages AI analysis and scientific methodologies to give investors the most comprehensive understanding of a company's ESG performance
- *Goldbaum* is a data provider for ESG and stock market data. Their intelligence is system-ready, and investors can make unlimited data API requests [?]
- *ISS ESG* is a leading provider of ESG data and climate-conscious solutions. Their data is used for asset and hedge fund management by global investors and corporates. ISS ESG's data makes investment in material assets more profitable and sustainable
- *Sense Folio* is an ESG data provider. They cover over 20,000 companies and deliver ESG score to investors on a daily basis, so that investment decisions can be based on real-time analytics
- *Miotech* offers stock exchange and ESG data for stock markets in China. Miotech monitors the ESG performance of over 800,000 companies across China, so investors can spot opportunities for alpha generation

Expanding body of work in ESG Finance will require new skills from quant analysts
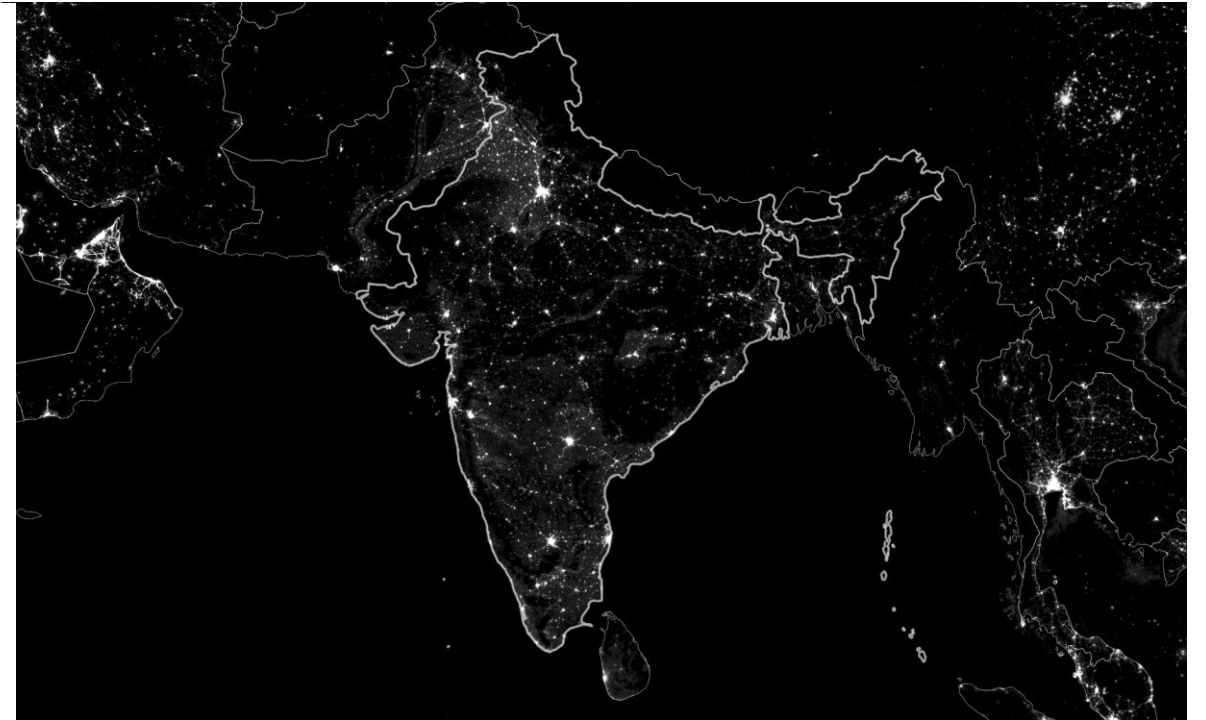
ESG Finance relies heavily unstructured or "alternative" data, because traditional structured sources are still very inconsistent. This requires specific new skills from practitioners

Example: natural language processing of corporate reports

- Translation Software
- Concept Learning

- Regular Expression
- Statistics

- Natural Language Generation
- Chatbots

- Automated Learning
- Non-Text Models

NLP ML DL AI

Example: satellite imagery used to monitor land and energy use

3.2 *Processing Satellite Data*

Example: Africa Night Light Index

An index now available on Bloomberg is a great example of alternative data (of the fourth kind – the satellite data) used directly for ESG evaluation purposes

Please refer to http://india.nightlights.io/#/about for an analysis of India night light using the same procedure

This section focuses on a very applied topic of processing satellite imagery, now (largely) freely available to an interested observer in large quantities. Given the growing importance of alternative data, a financial engineer will eventually be called upon to incorporate the information contained in imagery...

Source: http://india.nightlights.io/

Satellite Data

A number of currently operational satellites make their data available online, often for free (and there are many commercial data services catering to satellite data needs as well, of course)

The list of satellites/data sources below with their respective uses is compiled by Ben Heubl (https://gist.github.com/BenHeubl):

- **Sentinel-I**: Maritime and land monitoring, emergency response, climate change
- **Sentinel-II**: Land-cover maps, land-change detection maps, vegetation monitoring, monitoring of burnt areas
- **Sentinel-III**: Surface topography observations, ocean and land surface color observation and monitoring. The Sentinel-III OLCI instrument ensures the continuity of Envisat Meris
- **Sentinel-5P**: Monitoring the concentration of carbon monoxide (CO), nitrogen dioxide (NO_2), and ozone (O_3) in the air. Monitoring the UV aerosol index (AER_AI) and various geophysical parameters of clouds
- ESA's archive of **Landsat 5, 7 and 8**: vegetation monitoring, monitoring of land use, land cover maps, change monitoring. Provides global coverage of Landsat 8 - Envisat Meris and older data
- **Proba-V**: observation of land cover, vegetation growth, climate impact assessment, water resource management, agricultural monitoring and food security estimates, inland water resource monitoring and tracking of the spread of deserts and deforestation
- **MODIS**: Monitoring of land, clouds, ocean color at a global scale (by ESA)
- **GIBS**: Global image browser service with over 600 satellites made available by NASA

A very useful online page that provides a directory of relevant resources for satellite image analysis is https://gijn.org/resources-for-finding-and-using-satellite-images/

Spectral Bands

Sentinel 2 satellites have a Multi-Spectral Instrument (MSI) with 13 spectral bands that range from the visible range to the shortwave infrared (SWIR). Bands come in variable resolutions from 10 to 60 meters

Open access hub for Sentinel II data:
https://scihub.copernicus.eu/dhus/#/home

Natural Color (B4, B3, B2)

| Band	Resolution	Central Wavelength	Description
B1	60 m	443 nm	Ultra blue (Coastal and Aerosol)
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B8a	20 m	865 nm	Visible and Near Infrared (VNIR)
B9	60 m	940 nm	Short Wave Infrared (SWIR)
B10	60 m	1375 nm	Short Wave Infrared (SWIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

Normalized Difference Vegetation Index (NDVI)

A common type of analysis that is performed on satellite images that is very relevant to ESG is the calculation of a variety of brightness/color indexes in relevant geographic locations

Similar indexes are constructed for night lights, water surfaces, type of soil, and so on

Vegetation Index (B8-B4)/(B8+B4)
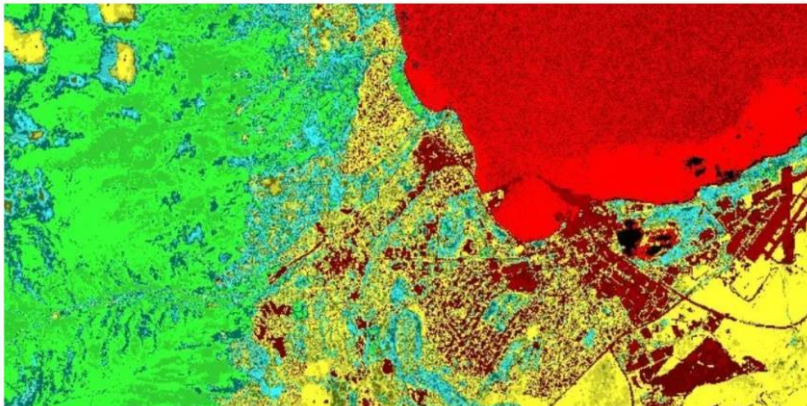
NDVI: Given a satellite image, one can calculate the "normalized difference vegetation index (NDVI)" for each point on the surface. The value of this index tells us (in quantifiable form) whether a particular point on the surface is covered by vegetation or not. NDVI is defined as

$$(b_8 - b_4) / (b_8 + b_4),$$

where b8 and b4 are intensities of the 8th ("near infrared" or NIR) and 4th (red) spectral bands respectively

Vegetation strongly reflects near-infrared and absorbs red light, so the NDVI index is good for quantifying the amount of vegetation. The higher the index value (see legend color bar on the right), the more likely the location is to be covered with forest

Similar Intensity-Based Indexes

- **Soil Adjusted Vegetation Index (SAVI)** corrects NVDI for local soil characteristics by introducing an L parameter:

$$SAVI = (1 + L) * (NIR - Red) / (NIR + Red + L)$$

- **Modified Normalized Difference Water Index (MNDWI)** is designed to enhance open water boundaries. Its primary application is the monitoring of the natural water reservoirs and the amount of water available. It is defined as the normalized difference between the Green band and NIR:

$$MNDWI = (Green - NIR) / (NIR + Green)$$

- **Visible Atmospherically Resistant Index (VARI)** is a variant of the index designed to mitigate illumination differences and atmospheric effects. It is defined as

$$VARI = (Green - Red) / (Green + Red - Blue)$$

- **Normalized Difference Moisture Index (NDMI)** describes the crops' water stress level and is calculated as the ratio between the difference and the sum of the refracted radiations in the near infrared and shortwave infrared (SWIR) bands:

$$NDMI = (NIR - SWIR) / (NIR + SWIR)$$

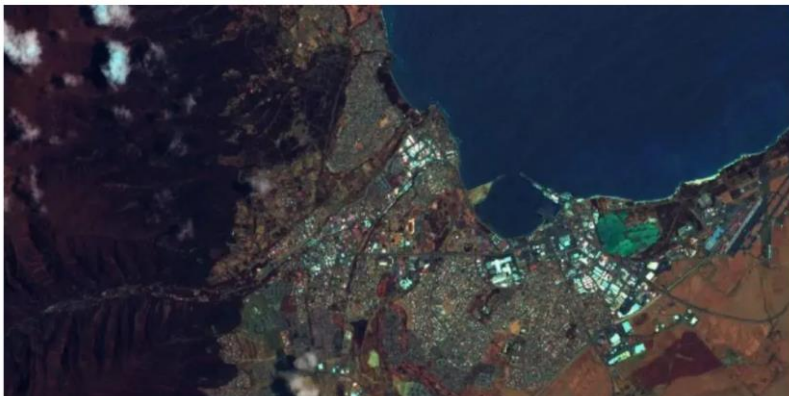# Intensity-Based Indexes for Geology

Of course, satellite/aerial imagery data have numerous applications beyond monitoring of the biosphere

For instance, similar intensity-based indexes can be useful in classifying soil type, which is an important predictor of fossil fuel and other mineral deposits

- Clay Minerals Ratio: ratio of SWIR and SWIR2
- Ferrous Minerals Ratio: ratio of SWIR and NIR (*SWIR / NIR*)

...are useful measures indicating the likely presence of respective mineral types

Geology (B12, B11, B2)

[Exercise: Quantitative Analysis of Satellite Images using Intensity-Based Indexes]
Rome data and parts of code courtesy of Syam Kakarla and Abdishakur Hassan

Session III Postface and Homework

Much of the "industrial grade" data processing conducted by alternative data providers relies on machine learning techniques where a neural network or another classification model is trained to recognize and find specific image patterns in large sets of satellite imagery. Deep learning is obviously a very broad (and deep!) topic of its own, which we do not have room to cover here. Obviously, by now there are many excellent textbooks dedicated to deep learning models and specifically image recognition

Define a triangular area in or around Sahara, download the sentinel II imagery using the open access hub, and show an NDVI index map for it in your Jupyter notebook

Please turn in either a Jupyter notebook or a script in a text file