

11th Biennial IFC Conference on “Post-pandemic landscape for central bank statistics”

BIS Basel, 25-26 August 2022

Microdata base for sustainability indicators (ESG) developed at the Banco de España¹

Borja Fernández Rosillo San Isidro,
Bank of Spain

¹ This presentation was prepared for the conference. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the event.

Microdata base for sustainability indicators (ESG) developed at the Banco de España

Analysing climate change data gaps

Borja Fernández-Rosillo San Isidro

Abstract

This paper focuses on the various limitations encountered and achievements made in the process of developing a microdata base for sustainability indicators for non-financial companies. After carefully researching ESG standards, consulting ESG experts, analysing regulatory obligations in Europe and conducting practical research, a list of the 39 most relevant ESG indicators was selected from those normally reported by companies. We have currently gathered more than 15,000 indicators (2019-2020) using an internally developed semi-automatic search application. Numerous limitations were identified in the process, in terms of different metrics, lack of information and lack of digital support for downloading, comparability difficulties and regulatory restrictions.

Keywords: ESG (Environmental, Social, Governance), data gaps, standards, climate change, sustainability, databases, metrics, digitalised, regulation.

JEL classification: Q5 (environmental economics).

Index

Analysing climate change data gaps.....	1
1. Introduction.....	3
2. Researching and establishing the relevant indicators	4
3. Extracting ESG information – transformation from unstructured to structured form	5
4. Data gaps and limitations in current ESG reporting.....	6
5. Improving data quality: quality control of sustainability indicators.....	11
6. Conclusions.....	13
Next steps.....	14
References.....	15
Annex 1 – Selection of indicators	15
Annex 2 – Dictionary of words associated with each indicator	16

1. Introduction

In recent years, awareness of social and environmental issues has been increasing, and consequently the demand for sustainability data has been growing exponentially. Although regulation and ESG reporting have been evolving and today there are numerous ESG requirements for companies, there is still a long way to go in terms of data availability, homogeneity, robustness and reliability. Despite its richness, the granular information available is still insufficient since:

It does not cover the entire population of companies, mainly because (i) current regulations exclude small and medium-sized enterprises from obligatory reporting requirements and (ii) companies belonging to a group are not obliged to report ESG indicators if the parent company of the group reports this information.

It is not homogeneous as there is no single definition of the indicators and metrics to be reported. Each company bases its reporting on a series of different standards with numerous ways of calculating and measuring the ESG indicators which makes comparison for the purposes of analysis difficult. Moreover, deriving from the first limitation, we have a mix of consolidated data (from companies which present their emissions for the entire group, i.e.) with individual data (from companies that directly report their emissions, i.e.). However, with the current regulatory advances, namely the Corporate Sustainability Reporting Directive (CSRD) and the European Sustainability Reporting Standards (ESRS) being developed in the European context by the European Financial Reporting Advisory Group (EFRAG) and the IFRS Sustainability Disclosure Standard being developed by the International Sustainability Standards Board (ISSB), as well as the Sustainability Accounting Standards Board (SASB) from the Value Reporting Foundation, the homogeneity limitation in relation to the indicators and metrics being reported is mostly being solved.

It is not digitalised, which makes it difficult to download data for processing. The ESG information is still not reported in a digital reporting language such as XBRL (Extensible Business Reporting Language). In the near term it is expected to be digitalised like the financial statements.

Taking into account the limitations mentioned above and bearing in mind the need to assess the impact of economic policy measures on climate change, as well as the role of the financial system in channelling investments towards environmentally sustainable activities, the Banco de España is interested in using the information available in order to generate ESG statistics. Consequently, and in line with the institution's objectives, the Statistics Department is working on the development of a microdata base for sustainability indicators for non-financial companies, the technical part of which has already been presented at the Irving Fisher Committee (IFC)¹ and, moreover, participating actively in three working groups relating to climate change².

¹ For further information regarding this web application prototype, see Koblents, Eugenia and Morales Alejandro, "Creation of a structured sustainability database from company reports: A web application prototype for information retrieval and storage", presented at the IFC.

² Working groups: (1) Workstream on exposure of financial institutions to climate-related physical risks from the STC Expert Group on Climate Change and Statistics / (2) Workstream on climate-related data gaps from the FSC Task Force on Climate Impact Assessment Analytics / (3) Workstream on ESG non-financial information reporting from the XBRL Spanish association.

The objective of this paper is to present the process developed in the Central Balance Sheet Data Office for obtaining, processing and analysing the ESG indicators of non-financial corporations.

2. Researching and establishing the relevant indicators

Before determining what indicators to look for, a rigorous analysis of the different aspects that needed to be taken into consideration was conducted, as follows:

Analysing the current regulatory obligations. In this part it was detected that, under the Non-Financial Reporting Directive (Directive 2014/95/EU that was transposed into Spanish law by Ley 11/2018), companies were obliged to present a report containing non-financial information. However, the precise indicators and the format for reporting them were not specified. Additionally, they were not reported digitally and they did not follow a structured form.

Researching the national and international ESG standards. In order to better understand the current indicators reported by companies and the technical aspects of their preparation different ESG standards were analysed. We focused on the GRI (Global Reporting Initiative) technical papers since most Spanish companies report in line with this standard.

Establishing a preliminary list of 120 indicators. First, lists of indicators were consulted, such as the one compiled by the Spanish Association of Accounting and Business Administration (AECA) and the one compiled by the private company INFORMA. In conjunction with the indicators reflected in the GRI standards, these consultations enabled a preliminary list of 120 indicators to be drawn up from which to select the most feasible and interesting ones.

Conducting a practical research exercise involving six listed companies. In order to select the indicators with the highest probability of being found and of greatest interest a practical exercise was performed by searching for the 120 indicators on the preliminary list in the non-financial reports of six listed companies published in the year 2019. Additionally, a survey was conducted among three experts (one from the rating and risk sustainability field, another from the sustainability finance field and the last one from the accounting and ESG reporting field) who categorised each indicator into one of three categories (low, medium and high), according to its potential interest. The result of this exercise, taking into consideration various aspects of each indicator (difficulty, interest, feasibility, etc.), was a list of 39 ESG indicators (Annex 1) to search for. Figure 1 and Chart 1 below depict the distribution of the 39 selected indicators by type (environmental, social or governance) and subtype (energy, water, etc.):

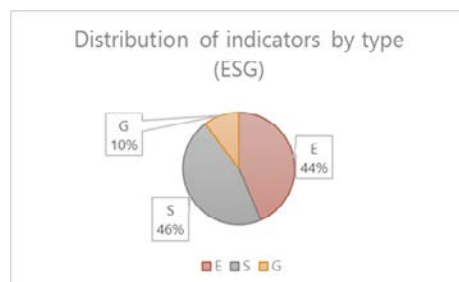


Figure 1. Distribution of the 39 selected indicators by type (ESG)

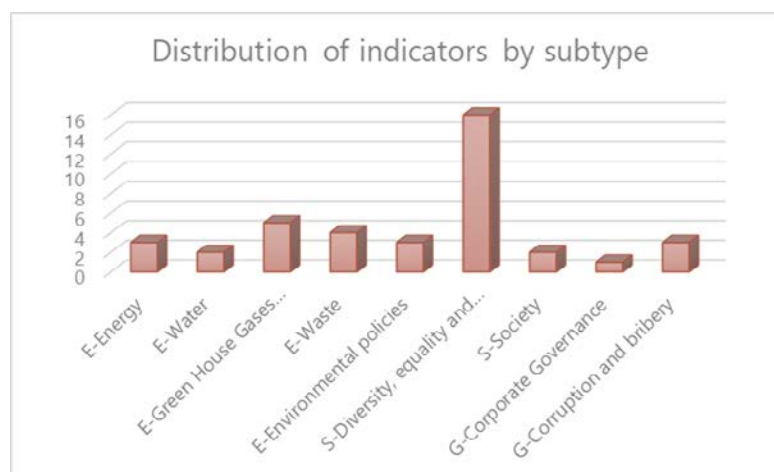


Chart 1. Distribution of the 39 selected indicators by subtype (energy, water, etc.)

3. Extracting ESG information – transformation from unstructured to structured form

Having established the indicators to look for it was necessary to address the process of retrieving them from non-financial reports. The main issue was the lack of homogeneity and digitalisation in the indicators reported. To solve this issue, and in close collaboration with data scientists of the Statistics Department, a semi-automatic search application was developed in order to obtain this data and transform it from an unstructured format to a structured database.³

In order to facilitate the gathering of this information the search process was based on a set of pre-defined terms for each sustainability indicator. The dictionary of ESG terms associated with each indicator was prepared on the basis of the practical exercise mentioned before involving six listed companies and the information supplied by the indicator descriptions obtained from the various standards consulted. The result was a preliminary list of words (see Annex 2) which would help the web application locate this information.

The first list of words was compiled without any experience of this process of extracting ESG indicators from non-financial reports. Now, having gathered approximately 15,000 indicators, and having saved the search terms for each indicator, there is a record of the degree of success of each list of words that enables the current ontology to be improved in order to increase the success rate. Natural Language Processing (NLP) and Artificial Intelligence (AI) tools enable the application to learn from current searches and automatically improve the ontology in order to increase the degree of success.

³ For further information regarding this web application prototype, see Koblenz, Eugenia and Alejandro Morales, "Creation of a structured sustainability database from company reports: A web application prototype for information retrieval and storage", presented at the IFC.

4. Data gaps and limitations in current ESG reporting

From the experience of this project there is enough evidence to confirm that most companies in the sample analysed are aware of ESG risks and reporting. However, a wide variety of limitations were found during the process.

Limitation 1 - Different metrics. During the process of extracting ESG information we found a wide variety of metrics for some indicators which were, at first, a clear barrier to direct comparison. However, in most cases it was possible to perform a simple transformation of the indicator into the homogeneous metric defined. Some examples for specific indicators are shown below:

Examples of “energy consumption within the organisation”

ENERGÍA	2015	2016	2017	2018	2019
Consumo total de Energía (MWh)	7.031.436	6.865.919	6.901.216	6.991.253	6.958.516
Electricidad (Mwh)	6.612.778	6.391.248	6.461.695	6.543.895	6.574.002

ENERGY	2015	2016	2017	2018	2019
Total energy consumption (MWh)	7,031,436	6,865,919	6,901,216	6,991,253	6,958,516
Electricity (Mwh)	6,612,778	6,391,248	6,461,695	6,543,895	6,574,002

Table 1. This table shows “energy consumption within the organisation” indicator for Telefonica (Spanish telecommunication company) in MWh

CONSUMO ENERGÉTICO INTERNO POR FUENTE PRIMARIA (TJ)*			
Tipo de combustible	2017	2018	2019
Carbón	244.764	221.079	81.527
Fuel óleo	58.205	53.313	47.755
Gasóleo	33.357	34.59	34.457
Gas natural	67.676	51.160	64.932
Uranio	280.139	254.926	279.042
Total consumo ENDESA	684.142	615.336	507.614

INTERNAL ENERGY CONSUMPTION BY PRIMARY SOURCE (TJ)*			
Fuel type	2017	2018	2019
Coal	244,764	221,079	81,527
Fuel oil	58,205	53,313	47,755
Diesel oil	33,357	34,590	34,457
Natural gas	6,768	51.160,000	64,932
Uranium	280,139	254,926	279,042
Total ENDESA consumption	684,142	615,336	507,614

Table 2. This table shows “energy consumption within the organisation” indicator for Endesa (Spanish energy company) in TJ

302-1 CONSUMO ENERGÉTICO DENTRO DE LA ORGANIZACIÓN				
		2017*	2018*	2019
Combustibles utilizados en fuentes Estacionarias y Móviles (total) (GJ)	Diésel	6.058.020	5.167.428	4.532.451
	Fuel	78.994	98.703	157.533
	Gasolina	472.599	289.117	586.315
	Gas Natural	3.039.568	260.542	304.364
	Carbón	390.225	570.558	361.701
	Queroseno	21.189	20.221	24.938
	Propano	18.467	27.732	22.793
	LPG	11.540	6.600	6.856
	TOTAL	10.090.602	6.440.901	5.996.951

302-1 ENERGY CONSUMPTION WITHIN THE ORGANISATION				
		2017*	2018*	2019
Fuel used in Stationary and Mobile sources (total) (GJ)	Diesel	6,058,020	5,167,428	4,532,451
	Oil	78,994	98,703	157,533
	Gasoline	472,599	289,117	586,315
	Natural gas	3,039,568	260,542	304,364
	Coal	390,225	570,558	361,701
	Kerosene	21,189	20,221	24,938
	Propane	18,467	27,732	22,793
	LPG	11,540	6,600	6,856
	TOTAL	10,090,602	6,440,901	5,996,951

Table 3. This table shows “energy consumption within the organisation” indicator for Ferrovial (Spanish construction and transport company) in GJ

The variety of metrics used by companies to present this information is clear. However, in these specific cases, where there was a possibility of converting all the indicators to the same metric, a preliminary analysis was carried out to see which metric was the most common one (in the case of “energy consumption within the organisation”, MWH) and then all the data stored in the database were converted into that metric in order to facilitate comparison and economic analysis. Hence, the variety of metrics does not prevent analysis of this type of information, but it does make the process longer and more difficult. Hopefully, future regulatory standards will specify a common metric in order to minimise this limitation.

Examples of “GHG emissions intensity”

Intensidad de las emisiones GEI	CO ₂	7,76	7,16	Kg CO ₂ e./prod.
GHG emissions intensity	CO ₂	7.76	7.16	Kg CO ₂ e/output

Table 4. This table shows the “GHG emissions intensity” indicator for Euskaltel (Spanish telecommunication company) in Kg CO₂e / output

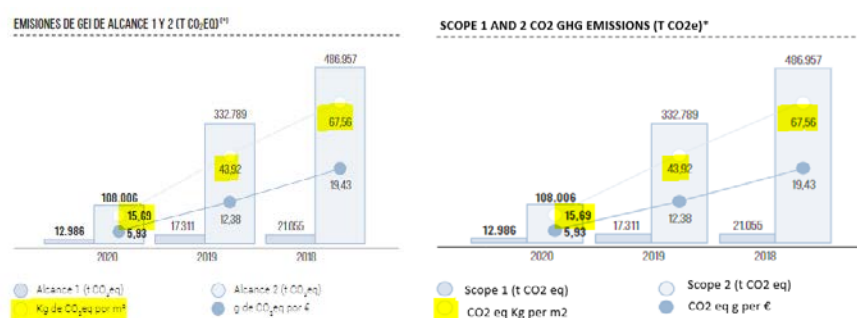


Chart 2. This table shows the “GHG emissions intensity” indicator for Inditex (Spanish retail company) in Kg CO₂e / m²

INTENSIDAD DE EMISIONES DE CO ₂ e			
T/CO ₂ e/millones de euros	2019	2018	2017
Total Grifols	64,8	66,6	69,3

ENERGY INTENSITY IN CO ₂ e			
TCO ₂ e/million euros	2019	2018	2017
Total Grifols	64.8	66.6	69.3

Table 5. This table shows the “GHG emissions intensity” indicator for Grifols (Spanish industrial company) in TnCO₂e / Millions €

In this second list of examples of regarding the “GHG emissions intensity” (Table 4 and 5 and chart 2) indicator there was a wide variety of ways in which companies presented this ratio as the current definition enables companies to use the business parameters they find most accurate (i.e. companies use in the denominator of the ratio total output, total sales, total employees, square metres of production plants, total km travelled, audio-visual production hours, etc.). Although some specific cases could be transformed (i.e. to use total employees or total sales, etc.), the numerator posed further difficulties as there was a variety of options to use (i.e. Scope 1, Scope 2 or Scope 1+2, among others). Consequently, in cases like this the data should only be taken into consideration for quality analysis, without attempting to make direct comparisons. In the case of this specific indicator it could be more beneficial to devise internally a ratio that can be calculated for all companies using the same parameters.

Limitation 2 - Changes in data over time. During the process it was not uncommon to see data change from one year to the next. There were a number of companies in the sample that changed data compilation criteria and, consequently, the number they had given for a specific indicator one year did not coincide with the number for the same indicator in the next year’s non-financial report. When an explanation was available in the non-financial report, they usually cited a change in the method of calculation. These changes are probably due to the novelty of this information and the scarcity of years of experience of collecting and preparing it. However, in these cases, the most up-to-date data were taken to be the relevant ones. Our intention is to closely follow the evolution of this information in order to see if it stabilises over time and if the number of changes falls significantly.

Distribución de las emisiones según alcances:

Emisiones ¹ (TCO ₂ eq)	2018*	2018 (comparable con 2019)	2019
Alcance 1	7.477	36.223	32.761
Alcance 2	949	949	1.159
Alcance 3	29.471	725	532
TOTAL	37.897	37.897	34.453

*Desglose reportado en EINF 2018

Distribution of emissions according to scope:

Emissions (TCO ₂ eq)	2018*	2018 (comparable with 2019)	2019
Scope 1	7,477	36,223	32,761
Scope 2	949	949	1,159
Scope 3	29,471	725	532
TOTAL	37,897	37,897	34,453

* Reported breakdown in 2018 non-financial statement

Table 6. This table shows the “Scope 1,2 and 3 GHG emissions” indicator for Grupo Ezentis SA (Spanish telecommunications and industrial company) in TnCO₂e.

In the example reflected above, the company presents in its 2019 non-financial report 2018 data that have been recalculated using a different method. This leads to a substantial difference between years for the same indicator. During this initial period of adaptation to ESG reporting, which companies are currently in, such changes in criteria and data will be common. However, after a reasonable time of compiling and processing non-financial information, these changes in criteria and data can be expected to be fewer and of less impact.

Limitation 3 - Lack of information. Some indicators were not found for some companies in the search process. Moreover, it is important to highlight that searches for some indicators from the preliminary selection list had a higher success rate than for others.

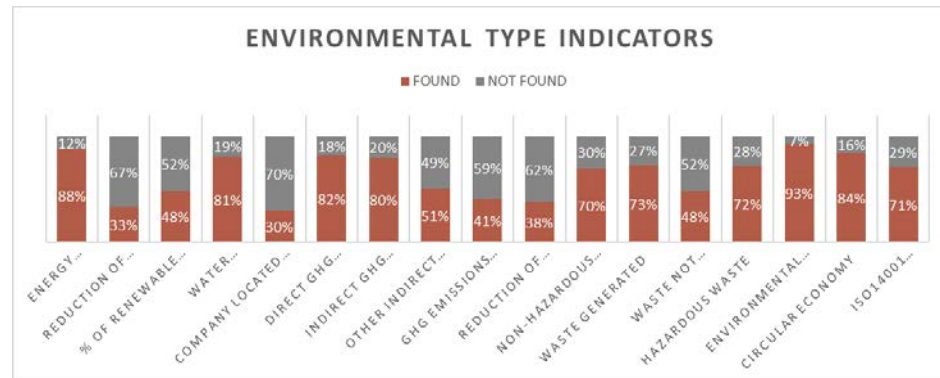


Chart 3. Distribution of found vs not found for environmental indicators (2019)

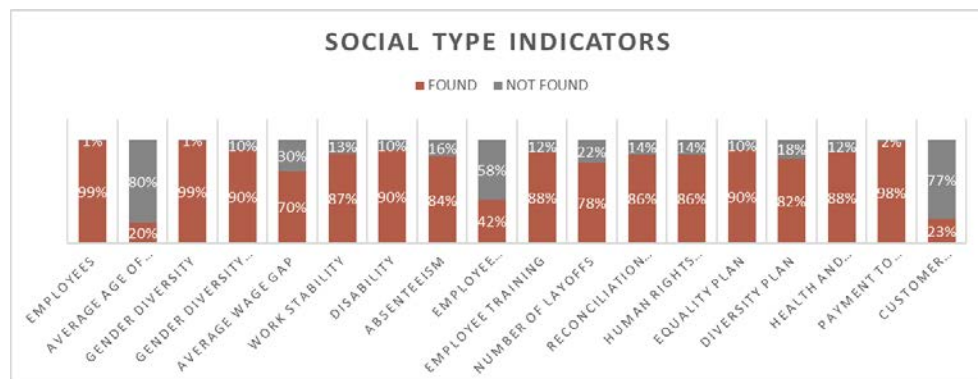


Chart 4. Distribution of found vs not found for social indicators (2019)

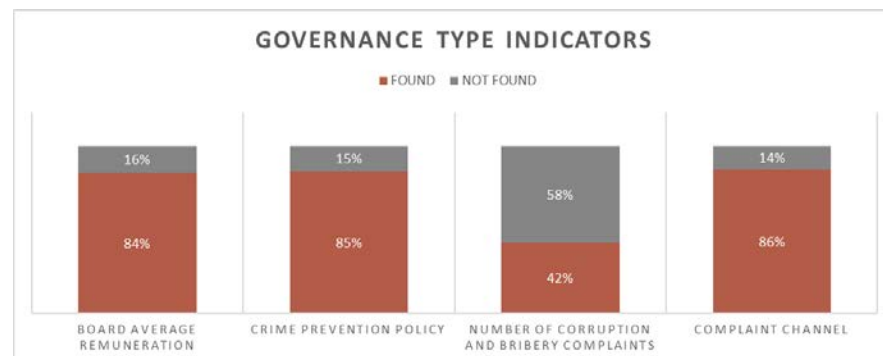


Chart 5. Distribution of found vs not found for governance indicators (2019)

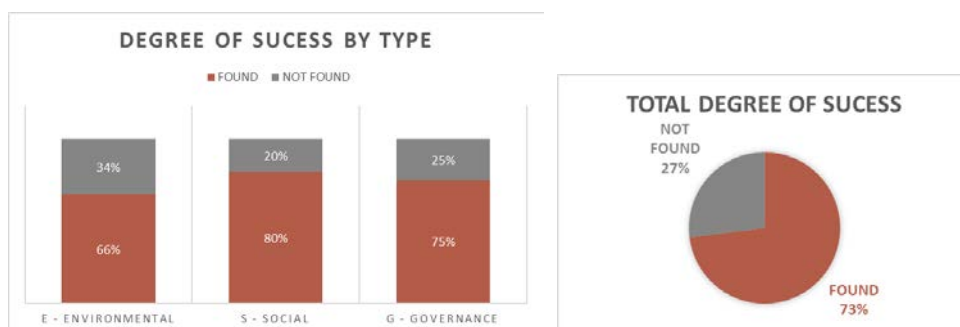


Chart 6. Distribution of found vs not found by type of indicator (ESG) and total (2019)

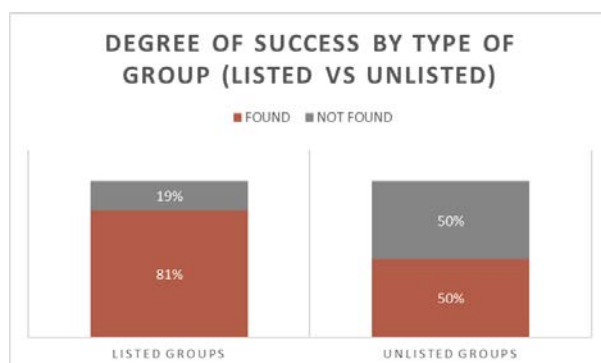


Chart 7. Distribution of found vs not found by type of company (listed vs unlisted) (2019)

The following conclusions drawn from the analysis presented above may be highlighted:

- Social indicators stand out as the type of indicator with the highest degree of success (80%) in their location, followed by governance (75%) and environmental (66%). This result was to be expected since social information (employees, diversity data, disability...) was being reported long before the current boom in environmental information.
- The overall success rate stands at 73%
- A substantial difference was observed between the degree of success in listed companies (81%) vs unlisted companies (50%). This is because listed companies have greater public exposure and stakeholder demand than unlisted companies.

Limitation 4 – Comparability difficulties. Although the main comparability issue is related to diverse metrics and can be relatively easily solved there are still some issues that generate difficulties in comparisons of this type of data such as:

Individual vs consolidated data. The current regulations exempt individual companies from presenting the non-financial report when the company reports the information at a consolidated level. Consequently, we have the inconvenience of gathering data which in some cases are to be found at a consolidated level and in others (less often) at an individual level.

Global data from international companies. Currently, many companies, especially the biggest, have an international presence all over the world. This affects the ESG information reported in their non-financial reports as it refers to the performance of the company or group worldwide, hindering the task of measuring national impacts (i.e. Scope 1 emissions in Spain). However, for a very small

percentage of companies and indicators, a breakdown is presented by country or economic region.

Different calculation methodologies. For some indicators and companies, the method of calculation varies so, although theoretically data could be comparable there is not enough evidence that the way they have obtained that data is through the same method or with similar criteria. A case in point are the Scope 2 emissions, which may be reported using a market approach or a location based approach. Consequently, if the company reports the method used there is no problem, but in the majority of the sample analysed the calculation method was not specified in the non-financial report.

Limitation 5 - Lack of digitalisation of the information presented. The information presented in the non-financial reports is still not digitalised and is reported in a wide variety of formats (i.e. charts, tables, etc.) which makes the process of locating it more difficult. However, the CSRD (Corporate Sustainability Reporting Directive) in progress includes the digitalisation and standardisation of reporting standards.

Limitation 6 - Lack of official ESG verification. Although there have been some advances in terms of ESG verification there is still no rigorous technical supervision of the information presented. Additionally, although the information presented could be checked, it is necessary to increase the amount of verification through the introduction of objective and technical measures that guarantee the quality and veracity of the data presented (i.e. that verify that the emissions a company reports are correctly calculated and real).

5. Improving data quality: quality control of sustainability indicators

The heterogeneity of the information and the novelty of having to analyse new types of information (e.g. greenhouse gas emissions and energy consumption) made it extremely important to conduct a rigorous and robust quality control assessment of the data. Consequently, each individual datum was reviewed, compared and contextualised, in terms of activity, in order to guarantee consistency in the database.

Moreover, during the process of analysis of each individual indicator the most common metric was identified in order to be able to define the standard metric in which to convert all the data to facilitate comparison. Additionally, and as a consequence of learning from this new information an interval analysis was conducted in order to establish the possible maximum and minimum values for each indicator. This exercise allowed the possible values of an indicator to be narrowed down and improved the process of searching for this information in future.

To increase the accuracy and quality of future uploadings of ESG data, individual guides were prepared for each of the indicators with the following fields:

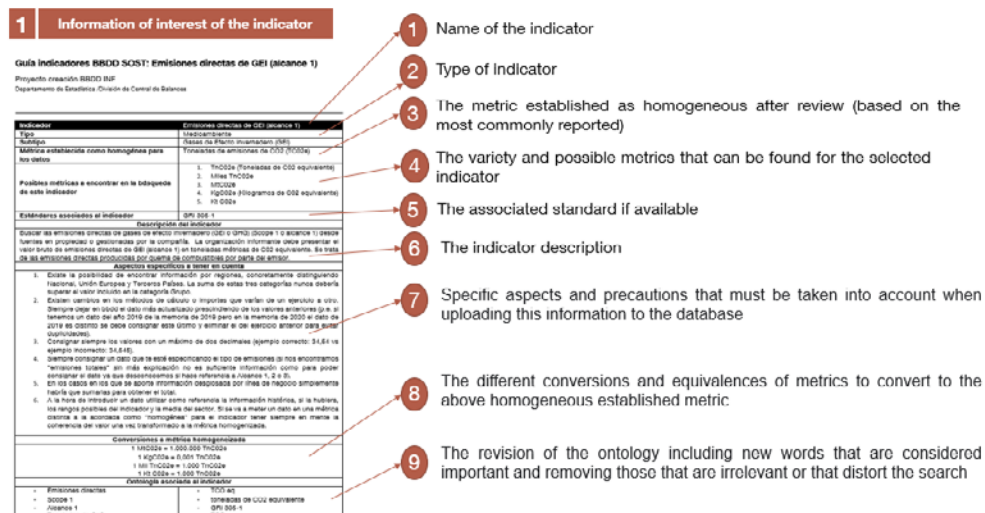


Figure 2. Example of the aspects covered in the first part (1-Information of interest on the indicator) of the guides prepared for each indicator

This first part (1-Information of interest on the indicator) of the guides relates to general information on the indicator in order to improve the process of locating the information. The specific aspects section contains information related to alternative ways of calculating the value or concrete aspects relevant to its interpretation. Additionally, the metrics sections help to establish if the data being uploaded is consistent in terms of value.

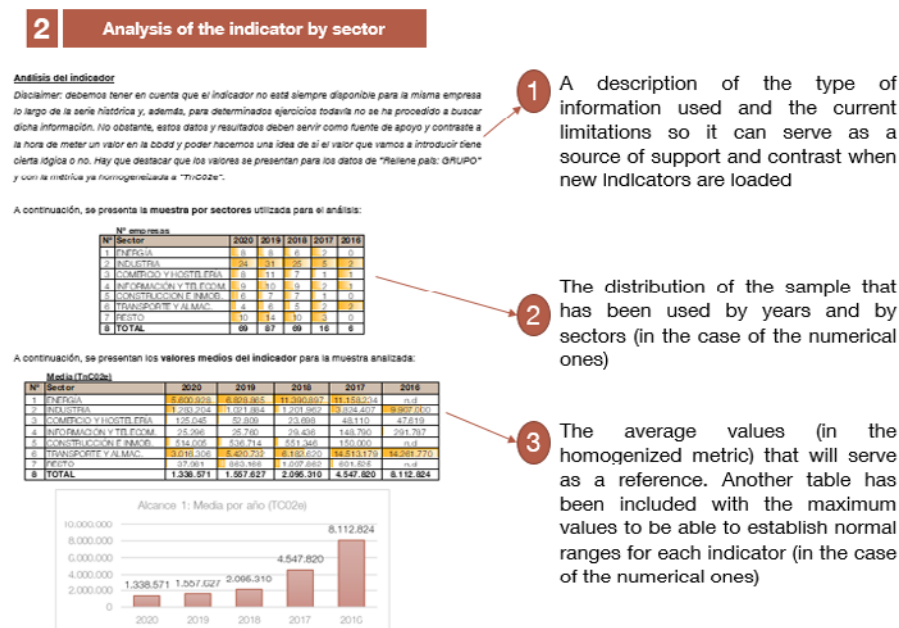


Figure 3. Example of the aspects covered in the second part (2-Analysis of the indicator by sector) of the guides prepared for each indicator

This second part (2-Analysis of the indicator by sector) of the guides relates to sample data for the indicator being analysed in order to provide a reference to establish if the new data being introduced into the database are consistent and realistic. This table allows the possible intervals of the indicator to be narrowed down and gives the average values by sector and year in order to minimise the potential number of future errors that can be introduced into the database.

1 Various examples of the way in which this information can be found in the non financial reports and that serves as a support element in the search for this type of information

3 Real examples in non-financial reports

Ejemplos de memorías (Scope 1)
A continuación, se muestran una serie de ejemplos en referencia al Scope 1 y la manera en la que las empresas lo están mostrando en sus memorías.

Ejemplos CGLA (Scope 1)

	2021	2020	2019	2018	2017
Emisiones de CO ₂	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de CH ₄	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de N ₂ O	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de HFC	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de PFC	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de SF ₆	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de CO ₂ equivalentes	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000

3.2 Cambios climáticos y energía
Total de emisiones de CO₂ equivalentes de la actividad de producción de corriente por país (normalizado)

	2021	2020	2019	2018	2017
Emisiones de CO ₂	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de CH ₄	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de N ₂ O	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de HFC	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de PFC	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de SF ₆	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de CO ₂ equivalentes	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000

3.3 Emisiones

	2021	2020	2019	2018	2017
Emisiones de CO ₂	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de CH ₄	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de N ₂ O	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de HFC	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de PFC	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de SF ₆	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000
Emisiones de CO ₂ equivalentes	1.740.000	1.740.000	1.740.000	1.740.000	1.740.000

Figure 4. Example of the aspects covered in the third part (3-Real examples on non-financial reports) of the guides elaborated for each indicator

This third and last part (3-Real examples in non-financial reports) of the guides relates to real examples of how this information can be presented in the non-financial reports in order to facilitate the location of this information.

6. Conclusions

Sustainability reporting is becoming increasingly relevant to measure companies impact on climate change. The numerous demands from the economy's stakeholders and the new green investment and financing decisions based on this information are key to the exponential growth in ESG reporting. However, despite current progress there is still a long way to go as there are still data gaps, heterogeneity in reporting, lack of digitalisation and verification and regulatory limitations.

The experience obtained in gathering ESG indicators from Spanish non-financial corporations' reports shows that listed companies present more and richer information than unlisted companies. Moreover, the ability to measure national ESG impacts is still a huge issue as multinational companies present information for the whole group and local companies either do not present ESG information, as they are exempted from doing so by the current regulations, or they belong to a group which presents consolidated data.

All in all, this project, in combination with the regulatory advances currently being made (CSRD) and the increasing awareness of reporting by companies, means that statisticians can be optimistic about ESG data availability in the near future.

Next steps

The achievements made and the experience obtained to date have provided the necessary catalyst to continue working on the provision of a robust ESG microdata base that can meet the various requirements of stakeholders and help minimise current ESG data gaps. Consequently, and bearing in mind the long-term goal of this project, the following lines of work are in the pipeline:

- Increasing the sample of companies from which information is currently extracted.
- Developing robust and solid quality control tests to minimise inconsistencies and errors in the data (automatic quality checks for future information uploads).
- Adapting the upcoming new ESG regulation to the microdata base.
- Analysing the possibility of increasing the number of indicators searched for.
- Refining the search process to increase the degree of success in terms of location and automation.

References

- [1] Koblents, Eugenia and Alejandro Morales. "Creation of a structured sustainability database from company reports: A web application prototype for information retrieval and storage" (2022)
- [2] Moreno, Ángel Iván, and Teresa Caminero. "Application of text mining to the analysis of climate-related disclosures." (2020).

Annex 1 – Selection of indicators

Selection of indicators on which sustainability information has been collected (39)

Type (E,S,G)	Subtype	Indicator
E	Energy	Energy consumption within the organisation
E	Energy	Reduction of energy consumption
E	Energy	% of renewable energy among the total
E	Water	Water consumption
E	Water	Company located in a stress area regarding water
E	Greenhouse gases (GHG)	Direct GHG emissions (scope 1)
E	Greenhouse gases (GHG)	Indirect GHG emissions when generating energy (scope 2)
E	Greenhouse gases (GHG)	Other indirect GHG emissions (scope 3)
E	Greenhouse gases (GHG)	GHG emissions intensity
E	Greenhouse gases (GHG)	Reduction of GHG emissions
E	Waste	Non-hazardous waste
E	Waste	Waste generated
E	Waste	Waste not destined for disposal
E	Waste	Hazardous waste
E	Environmental policies	Environmental policy
E	Environmental policies	Circular economy
E	Environmental policies	ISO14001 regulatory compliance
S	Diversity, equality and well-being of staff	Employees
S	Diversity, equality and well-being of staff	Average age of employees
S	Diversity, equality and well-being of staff	Gender diversity
S	Diversity, equality and well-being of staff	Gender diversity on the board
S	Diversity, equality and well-being of staff	Average wage gap
S	Diversity, equality and well-being of staff	Work stability
S	Diversity, equality and well-being of staff	Disability
S	Diversity, equality and well-being of staff	Absenteeism

Type (E,S,G)	Subtype	Indicator
S	Diversity, equality and well-being of staff	Employee rotation
S	Diversity, equality and well-being of staff	Employee training
S	Diversity, equality and well-being of staff	Number of layoffs
S	Diversity, equality and well-being of staff	Reconciliation measures
S	Diversity, equality and well-being of staff	Human rights policy
S	Diversity, equality and well-being of staff	Equality plan
S	Diversity, equality and well-being of staff	Diversity plan
S	Diversity, equality and well-being of staff	Health and safety policy
S	Society	Payment to suppliers
S	Society	Customer satisfaction level
G	Corporate governance	Average board remuneration
G	Corruption and bribery	Crime prevention policy
G	Corruption and bribery	Number of corruption and bribery complaints
G	Corruption and bribery	Complaint channel

Annex 2 – Dictionary of words associated with each indicator

Type (E,S,G)	Indicator	Dictionary of words (original queries in Spanish)	Dictionary of words (English)
E	Energy consumption within the organisation	Consumo de energía energético eléctrico electricidad dentro de la organización KWh, MWh, GWh, TWh, kJ, MJ, GJ, TJ, PJ "302-1" dato total vatios julios	Consumption of energy electrical energy electricity within the organization KWh, MWh, GWh, TWh, kJ, MJ, GJ, TJ, PJ "302-1" total data watts joules
E	Reduction of energy consumption	Reducción del consumo energético Energía KWh, MWh, GWh, TWh, kJ, MJ, GJ, TJ, PJ "302-4" dato total evitar descender julios vatios	Reduction of energy consumption Energy KWh, MWh, GWh, TWh, kJ, MJ, GJ, TJ, PJ "302-4" total data avoid falling joules watts
E	% of renewable energy among the total	porcentaje energía renovable total % ratio fuentes TJ consumo energético combustible fósil verde KWh GWh origen "menores emisiones" alternativa eólica solar 302 "no convencionales" KWh, MWh, GWh, TWh, kJ, MJ, GJ, TJ, PJ	percentage of total renewable energy % ratio sources TJ energy consumption green fossil fuel KWh GWh origin "lower emissions" solar wind alternative 302 "non-conventional" kWh, MWh, GWh, TWh, kJ, MJ, GJ, TJ, PJ
E	Water consumption	"Consumo de agua" "uso de agua" "gasto de agua" "303-5" volumen caudal total dato hm m3 Hm3 megalitros litros millones toneladas masa captación dulce	"Water consumption" "water use" "water expenditure" "303-5" volume total flow data hm m3 hm3 megalitres litres million tonnes fresh catchment mass

Type (E,S,G)	Indicator	Dictionary of words (original queries in Spanish)	Dictionary of words (English)
E	Company located in a water stress area	Estrés hídrico localización zona área agua escasez recursos hídricos H20 303	Water stress location area water scarcity water resources H20 303
E	Company located in a water stress area	Estrés hídrico localización zona área agua escasez recursos hídricos H20 303	Water stress location area water scarcity water resources H20 303
E	Direct GHG emissions (scope 1)	Emisiones directas de GEI "alcance 1" "scope 1" "ambito 1" "Gases de Efecto Invernadero" Mt Tn Tm CO2* "305-1" miles millones toneladas dato total kt calcul*	Direct GHG emissions "scope 1" "scope 1" "scope 1" "Greenhouse Gases" Mt Tn Tm CO2* "305-1" billion tonnes total data kt calcul*
E	Indirect GHG emissions when generating energy (scope 2)	Emisiones indirectas de GEI genera* energía energético "alcance 2" "scope 2" "ambito 2" "Gases de Efecto Invernadero" Mt Tn Tm CO2* "305-2" miles millones	Indirect GHG emissions generated* energy "scope 2" "scope 2" "scope 2" "Greenhouse Gases" Mt Tn Tm CO2* "305-2" billions
E	Other indirect GHG emissions (scope 3)	Otras emisiones indirectas de GEI genera* energía energético "alcance 3" "scope 3" "ambito 3" "Gases de Efecto Invernadero" Mt Tn Tm CO2* "305-3" miles millones toneladas eléctrico electricidad dato total kt calcul* TCO eq equivalente tCO* dióxido de carbono GHG	Other indirect GHG emissions generated* energy energy "scope 3" "scope 3" "scope 3" "Greenhouse Gases" Mt Tn Tm CO2* "305-3" billion tonnes electric electricity total data kt calcul* TCO eq equivalent tCO* carbon dioxide GHG
E	GHG emissions intensity	Intensidad de las emisiones de GEI "Gases de Efecto Invernadero" Ratio "305-4" dato tCO* CO2* kg kilogramo carbono TJ intensidad energética MtCO2e GHG	Intensity of GHG emissions "Greenhouse Gases" Ratio "305-4" data tCO* CO2* kg kilogram carbon TJ energy intensity MtCO2e GHG
E	Reduction of GHG emissions	Reducción de las emisiones de GEI objetivo evitar "Gases de Efecto Invernadero" Mt Tn CO2* "305-5" miles millones toneladas dato TCO* eq equivalente GHG	Reduction of GHG emissions objective to avoid "Greenhouse Gases" Mt Tn CO2* "305-5" billions of tonnes TCO data* GHG equivalent eq
E	Non-hazardous waste	Residuos generados "no peligrosos" reutilización reciclaje recuperación incineración compostaje toneladas Tn "306-2"	Waste generated "non-hazardous" reuse recycling recovery incineration composting tonnes Tn "306-2"
E	Waste generated	Residuos generados peso total toneladas Tn "306-3"	Waste generated total weight tonnes Tn "306-3"
E	Waste not destined for disposal	Residuos generados no destinados a eliminación Tn Toneladas composición "306-4"	Waste generated not intended for disposal Tn Tonnes composition "306-4"
E	Hazardous waste	Residuos generados peligrosos Tn Toneladas 306	Hazardous waste generated Tn Tonnes 306
E	Environmental Policy	Medio Ambiente Política medioambiental Futuro limpio de emisiones Estrategia climática menor impacto productos bajos 30X	Environment Environmental policy Clean future of emissions Climate strategy less impact low products 30X
E	Circular economy	Proyectos de "economía circular" Reducción del uso de "materias primas" Ecología Circularidad Ecodiseño	"Circular economy" projects Reduction in the use of "raw materials" Ecology Circularity Ecodesign alternatives Efficiency in processes Innovation

Type (E,S,G)	Indicator	Dictionary of words (original queries in Spanish)	Dictionary of words (English)
alternativas Eficiencia en procesos Innovación			
E	ISO14001 regulatory compliance	ISO 14001 cumplimiento normativa Estándar en gestión ambiental Responsabilidades medioambientales reducir impacto corporativa Requerimientos regulatorios	ISO 14001 regulatory compliance Environmental management standard Environmental responsibilities reduce corporate impact Regulatory requirements
S	Employees	Número de empleados Mujeres Hombres Distribución plantilla Contrato laboral Fijos Temporales "102-8"	Number of employees Women Men Workforce distribution Employment contract Permanent Temporary "102-8"
S	Average age of employees	edad media de la plantilla años empleo empleados	average age of the workforce years of employment employees
S	Gender diversity	Nº de hombre* mujer* total empleadas Empleo Distribución por sexo género igualdad laboral integración "102-8"	No. of men* women* total employees Employment Distribution by sex gender labour equality integration "102-8"
S	Gender diversity on the board	distribución por sexo "102-8" Consejo de administración mujer* Composición Compromiso igualdad Consejeros	distribution by sex "102-8" Board of Directors female* Composition Commitment to equality Directors
S	Average wage gap	Brecha salarial Sueldos por sexo Igualdad de oportunidades Retribución Paridad Políticas de género a favor de hombre* mujer* equidad remuneración equitativa	Salary gap Salaries by sex Equal opportunities Remuneration Parity Gender policies in favour of men* women* equity equal pay
S	Work stability	Empleados por contrato laboral "102-8" Permanente Temporal Indefinidos Personal fijo tipo Tipología laborales	Employees by employment contract "102-8" Permanent Temporary Indefinite Permanent staff type Type of work
S	Disability	Discapacidad "405-1" Empleados Personal discapacitado grado Grupos vulnerables integración social inserción laboral igualdad de oportunidades accesibilidad exclusión diversidad discriminación respetuoso	Disability "405-1" Employees Disabled staff degree Vulnerable groups social integration job placement equal opportunities accessibility exclusion diversity discrimination respectful
S	Absenteeism	absentismo días puesto de trabajo ausencia ausentismo perdidos laborables laborales total horas jornadas	absenteeism days job absence absenteeism lost working days total hours working hours
S	Employee rotation	rotación de empleados abandono de puesto de trabajo fin relación laboral número personas abandonan "404-1" plantilla cese voluntario	employee turnover job abandonment termination of employment relationship number of people leaving "404-1" template voluntary resignation
S	Employee training	total horas de formación cursos empleado trabajador % año	total hours of training courses employee worker % year
S	Number of layoffs	Número de despidos empleados despedidos extinción rescisiones fin relación laboral cese indemnización terminación rescindir contrato trabajo	Number of layoffs dismissed employees termination terminations end of employment relationship cessation compensation termination terminate employment contract
S	Reconciliation measures	Conciliación bienestar empleados medidas de laboral políticas sociales reducir estrés emocional mental vida privada familiar equilibrio hijos mayores discapacitados compaginar conciliar	Reconciliation well-being employees labour measures social policies reduce mental emotional stress private family life balance older children disabled reconcile

Type (E,S,G)	Indicator	Dictionary of words (original queries in Spanish)	Dictionary of words (English)
S	Human rights policy	Compromiso social política de derechos humanos 412 vulneración abusos vulnerable responsabilidad	Social commitment human rights policy 412 violation abuses vulnerable responsibility
S	Equality plan	Compromiso social plan política de igualdad oportunidades "102-8" "406-1" "404-2" inclusión discriminación raza sexo color religión diversidad género mujer*	Social commitment equal opportunities policy plan "102-8" "406-1" "404-2" inclusion discrimination race sex colour religion diversity gender woman*
S	Diversity plan	Políticas de inclusión Igualdad de derechos LGBTI LGBTI Plan de diversidad plantilla 405 origen étnico orientación identidad sexual sin prejuicios	Inclusion policies Equal rights LGBTI LGBTI Diversity plan 405 template ethnic origin orientation sexual identity without prejudice
S	Health and safety policy	Salud y seguridad política de asistencia sanitaria seguros médicos beneficios sociales planes de ayuda beneficios sociales 403 bienestar de los empleados	Health and safety health care policy medical insurance social benefits assistance plans social benefits 403 employee welfare
S	Payment to suppliers	Periodo medio pago a proveedores días PMP cadena de suministro grupos de interés	Average payment period to suppliers days PMP supply chain stakeholders
S	Customer satisfaction level	Nivel de satisfacción cliente servicio postventa trato experiencia compra usuario grado "102-43" GRI "102-44" insatisfacción grupos interés	Level of customer satisfaction after-sales service treatment user purchase experience grade "102-43" GRI "102-44" dissatisfaction with interest groups
G	Board average remuneration	Remuneración media percibida salario del consejo retribución del consejo administración 405 miles anual euros variable especie monetaria fija a largo plazo	Average remuneration received salary of the board remuneration of the board of directors 405 thousand euros per year variable long-term fixed monetary species
G	Crime prevention policy	Modelo de Prevención de Delitos Certificado AENOR Sistema de gestión de compliance penal antisoborno 205 Anticorrupción UNE-ISO 37001 UNE 19601 soborno	Crime Prevention Model AENOR Certificate Anti-bribery criminal compliance management system 205 Anti-corruption UNE-ISO 37001 UNE 19601 bribery
G	Number of corruption and bribery complaints	Casos de corrupción soborno políticas antisoborno nº "205-3" fraude prevenir luchar sanción indebido denuncia	Corruption cases bribery anti-bribery policies number "205-3" fraud prevent fight sanction improper complaint
G	Complaint channel	Canal de denuncias Ética y Cumplimiento sistemas de comunicación comunicar fraud* empleados transparencia ilegal irregular* conducta Anonimato Confidencialidad 205 de forma confidencial y anónima	Ethics and Compliance Whistleblowing Channel Communication Systems Communicate Fraud* Employees Transparency Illegal Irregular* Conduct Anonymity Confidentiality 205 Confidentially and Anonymously

MICRODATA BASE FOR SUSTAINABILITY INDICATORS (ESG) DEVELOPED AT THE BANCO DE ESPAÑA - ANALYSING CLIMATE CHANGE DATA GAPS

Borja Fernández-Rosillo San Isidro

Central Balance Sheet Data Office
Statistics Department

11TH IFC CONFERENCE ON “POST-PANDEMIC LANDSCAPE FOR CENTRAL BANKS STATISTICS”

BASEL
25/08/2022



INDEX

1. Introduction and context
2. Researching and establishing the relevant indicators
3. Extracting ESG information: from unstructured to structured form
4. Data gaps and limitations
5. Improving data quality
6. Conclusions

1. INTRODUCTION AND CONTEXT

Why does this need arise? What are the main challenges?



In recent years, awareness of social and environmental issues has been increasing, and consequently the demand for sustainability data has been growing exponentially.



- 1 Measure the exposure of the Spanish Economy to climate change at a disaggregated level
- 2 Analyse the implications of climate change and transition to a more sustainable economy
- 3 Evaluate the impact of economic policy measures on climate change
- 4 Facilitate the channelling of investment towards environmentally friendly activities



NEED FOR A GRANULAR DATABASE ON ESG INFORMATION



It is essential to increase the quantity, quality and harmonization of Environmental information

Despite its richness, the granular information available is still insufficient since:

1. It does not cover the entire population

2. It is not homogeneous

3. It is not digitalised

1. INTRODUCTION AND CONTEXT

Most limiting aspects in current and future regulations

MAIN CURRENT LIMITATIONS

Spanish Law 11/2018 on non-financial information and diversity

CSRD (Corporate Sustainability Reporting Directive) – in progress

1

ONLY GROUPS? EXEMPTION FOR SUBSIDIARIES



2

HETEROGENEITY OF REPORTED INDICATORS



EFRA - **ESRS** (European Sustainability Reporting Standards)

3

LACK OF DIGITALIZATION



Information available in XBRL format

4

LIMITED POPULATION (SMALL AND MEDIUM OUTSIDE)



TWO IMPORTANT LIMITATIONS REMAIN WITH THE NEW CSRD

2. RESEARCHING AND ESTABLISHING THE RELEVANT INDICATORS

Main aspects taken into consideration

1. Analysing the current regulatory obligations

- Non-Financial Reporting Directive ([Directive 2014/95/EU](#) that was transposed into Spanish law by [Ley 11/2018](#))
- In the future CSRD will be transposed and will update Ley 11/2018

2. Researching the national and international ESG standards



3. Establishing a preliminary list of 124 indicators

List of preliminary indicators

TOTAL

E-Environmental
S-Social
G-Governance

124
62
44
18

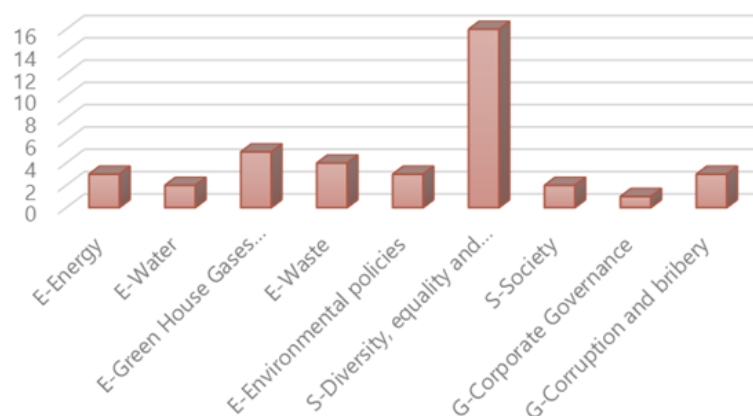
4. Conducting a practical research exercise involving six listed companies



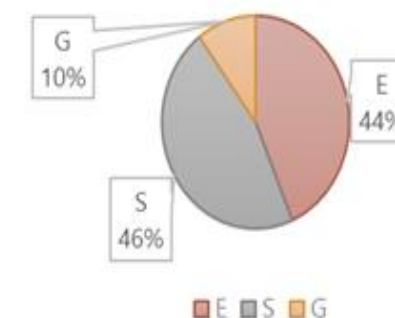
5. List of 39 ESG indicators to search for

$$\boxed{E} + \boxed{S} + \boxed{G} = \boxed{39}$$

Distribution of indicators by subtype

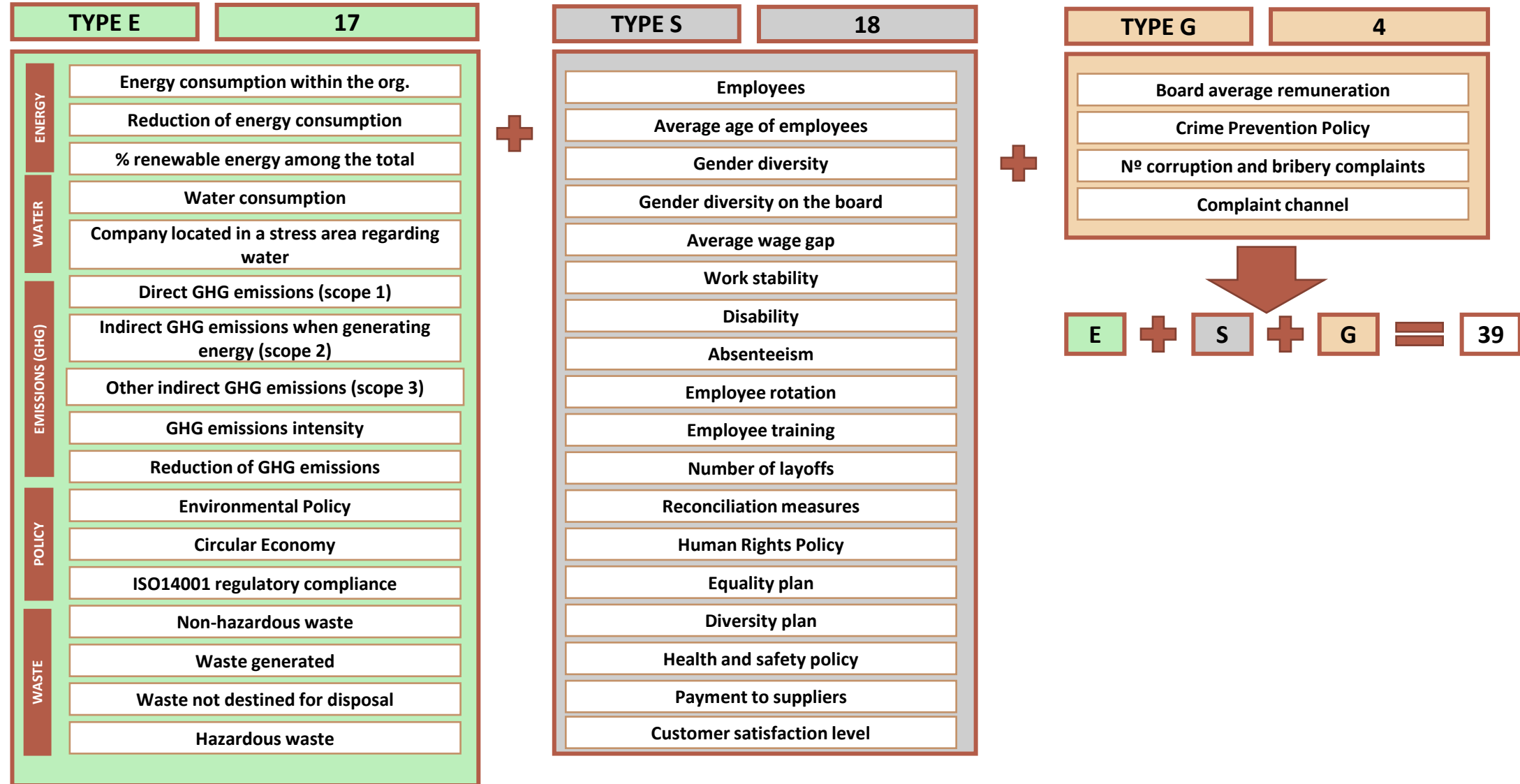


Distribution of indicators by type (ESG)



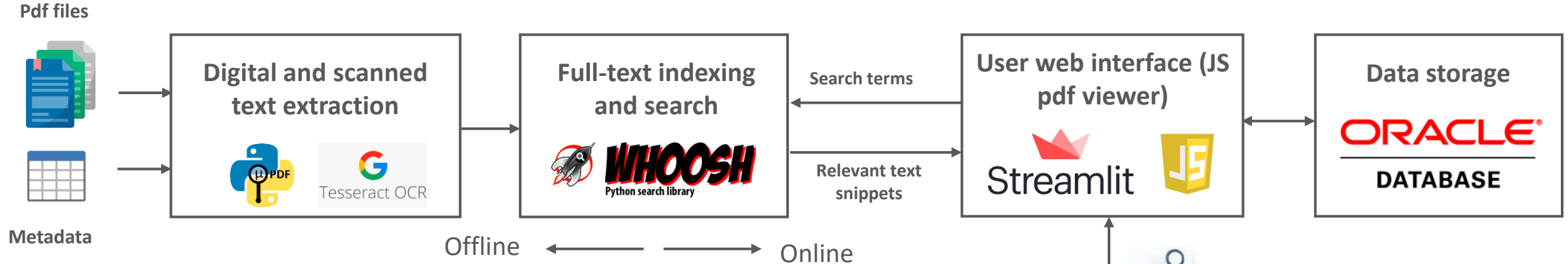
2. RESEARCHING AND ESTABLISHING THE RELEVANT INDICATORS

List of indicators for searching



3. EXTRACTING ESG INFORMATION: FROM UNSTRUCTURED TO STRUCTURED FORM

Prototype developed with AI (Artificial Intelligence) support*



- ❑ **Preliminary dictionary of ESG search terms:** helps the semi-automatic search application locate this information.
- ❑ **Context information (exact location and paragraph)** is saved for each indicator.
- ❑ This **labelled data** enables to automatically optimize search terms using **Machine Learning (ML)** and **Natural Language Processing (NLP)**. An strategy has been designed to propose, evaluate, compare and optimize queries.

The screenshot shows the web application interface. On the left, there's a sidebar with filters for 'Usuario', 'Año', 'Prioridad', 'Empresa', and 'NIF empresa'. The main area displays search results for 'Consumo de agua'. A table shows the 'Consumo de Agua (m3)' for the years 2020 and 2019, categorized by 'Extracción de agua', 'Vertido de agua', 'Agua vendida', and 'Total Consumo de agua (m3)'.

Consumo de Agua (m3)	2020	2019
Extracción de agua	37.737.355	20.677.055
Vertido de agua	2.663.210	2.199.031
Agua vendida	3.168	0
Total Consumo de agua (m3)	35.070.977	18.478.024

(*) For further information regarding this web application prototype, see Koblents, Eugenia and Alejandro Morales, "Creation of a structured sustainability database from company reports: A web application prototype for information retrieval and storage", presented at the IFC.

4. DATA GAPS AND LIMITATIONS

ESG data (1/2)

1

DIFFERENT METRICS

- **Wide variety** of metrics for some indicators
- Barrier to direct comparison
- However, in most cases it was possible to perform a simple **transformation** to a homogeneous metric defined
- Future regulatory standards will **specify a common metric** in order to minimise this limitation

2

CHANGES IN DATA OVER TIME

- During the process it was not uncommon to see **data change** from one year to the next
- These changes are probably due to the **novelty** of this information and the **scarcity of years of experience**
- We will closely follow the evolution of this information in order to see if it stabilises over time

3

LACK OF INFORMATION

- A substantial difference was observed between the degree of success in **listed companies** (81%) vs **unlisted companies** (50%).
- The overall success rate stands at 73%
- **Social indicators** stand out as the type of indicator with the **highest degree of success** (80%) in their location.



ENERGY	2015	2016	2017	2018	2019
Total energy consumption (MWh)	7,031,436	6,865,919	6,901,216	6,991,253	6,958,516
Electricity (MWh)	6,612,778	6,391,248	6,461,695	6,543,895	6,574,002



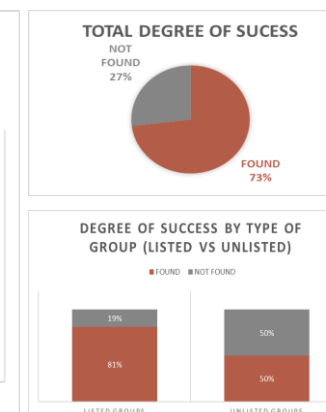
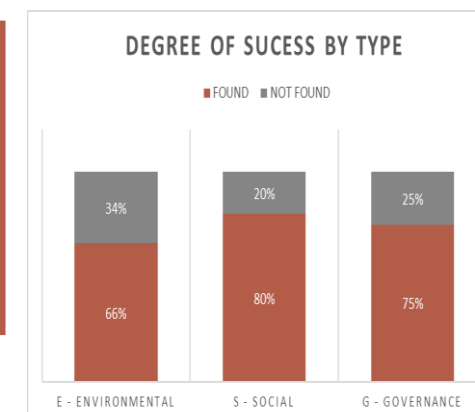
INTERNAL ENERGY CONSUMPTION BY PRIMARY SOURCE (TJ)*			
Fuel type	2017	2018	2019
Total ENDESA consumption	684,142	615,336	507,614

Distribution of emissions according to scope:



Emissions (TCO2eq)	2018*	2018 (comparable with 2019)	2019
Scope 1	7,477	36,223	32,761
Scope 2	949	949	1,159
Scope 3	29,471	725	532
TOTAL	37,897	37,897	34,453

* Reported breakdown in 2018 non-financial statement



4. DATA GAPS AND LIMITATIONS

ESG data (2/2)

4

COMPARABILITY DIFFICULTIES

- Due to **diverse metrics** (easily solution in most cases) and **different calculation methodologies**
- **Individual vs consolidated data**. Current regulatory exemptions provokes some inconvenience
- **Global data from international companies** makes it difficult to measure national impacts

5

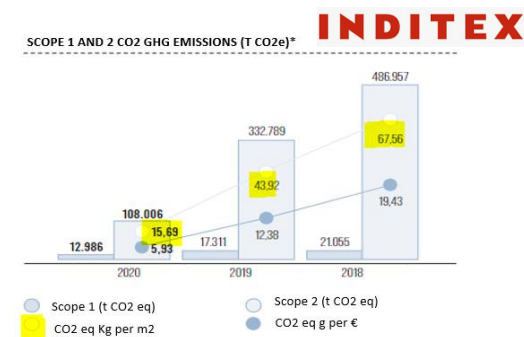
LACK OF DIGITALISATION OF THE INFORMATION PRESENTED

- Information presented in the non-financial reports is **still not digitalised**
- However, the **CSRD** (Corporate Sustainability Reporting Directive) **includes the digitalisation** and standardisation of reporting standards

6

LACK OF ESG VERIFICATION

- **No rigorous technical supervision** of the information presented
- Introduction of **objective and technical** measures that guarantee the quality and veracity of the data presented



GRIFOLS

ENERGY INTENSITY IN CO2e

TCO2e/million euros	2019	2018	2017
Total Grifols	64.8	66.6	69.3

ferrovial

302-1 ENERGY CONSUMPTION WITHIN THE ORGANISATION

	2017*	2018*	2019
Diesel	6,058,020	5,167,428	4,532,451
Oil	78,994	98,703	157,533
Gasoline	472,599	289,117	586,315
Natural gas	3,039,568	260,542	304,364
Coal	390,225	570,558	361,701
Kerosene	21,189	20,221	24,938
Propane	18,467	27,732	22,793
LPG	11,540	6,800	6,856
TOTAL	10,090,602	6,440,901	5,996,951

Fuel used in Stationary and Mobile sources (total) (GJ)

euskaltel

GHG emissions intensity	CO2	7.76	7.16	Kg CO2 e/output
-------------------------	-----	------	------	-----------------

Quality control of sustainability indicators

The heterogeneity of the information and the novelty of having to analyse new types of information (e.g. greenhouse gas emissions and energy consumption) made it extremely important to conduct a rigorous and robust quality control assessment of the data → **BANK OF SPAIN ESG GUIDES FOR INTERNAL USE***

1 Information of interest on the indicator

[illegible]

This **first part** of the guides relates to general information on the indicator in order to improve the process of locating the information.

2 Analysis of the indicator by sector

[illegible]

This **second part** of the guides relates to sample data for the indicator being analysed in order to provide a reference to establish if the new data being introduced into the database are consistent and realistic.

3 Real examples in non-financial reports

1

Various examples of the way in which this information can be found in the non financial reports and that serves as a support element in the search for this type of information

The screenshot shows a financial statement titled 'Financiamento' (Financing). It includes a table with columns for 'Financiamento', 'Financiamento de exploração', 'Financiamento de outros recursos', and 'Financiamento de recursos próprios'. The table lists various financial metrics and their values.

This **third part** of the guides relates to real examples of how this information can be presented in the non-financial reports in order to facilitate the location of this information.

(*) To increase the accuracy and quality of future uploadings of ESG data, individual guides were prepared for each of the 39 indicators. These guides are only available for Bank of Spain technical employees involved in the process of gathering ESG data.

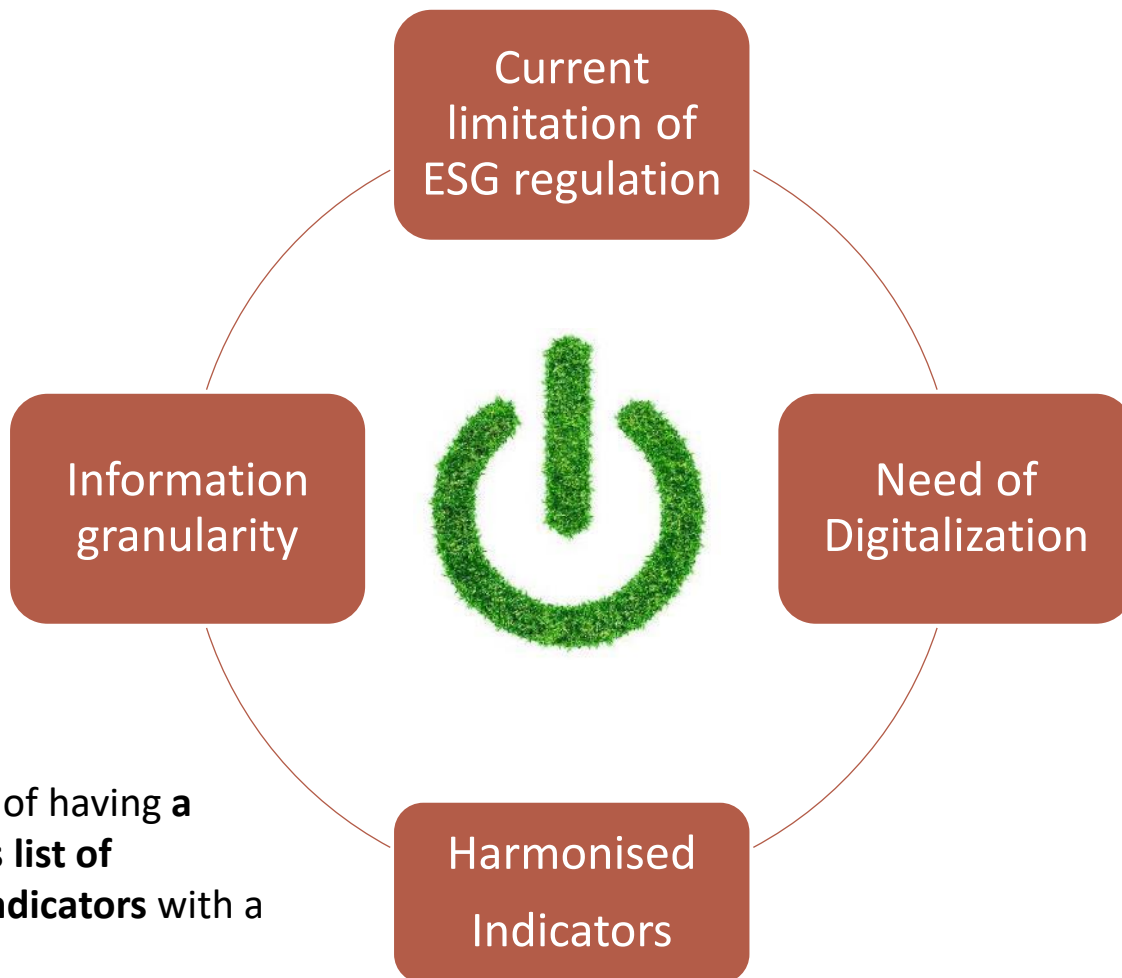
6. CONCLUSIONS

Conclusions and challenges

- ✓ **Real data shows the limitation of current ESG regulation** (data gaps, heterogeneity and variability in reporting, comparability difficulties...)

- ✓ **Information of groups is not enough.** There is a need for information from individual companies.

- ✓ The relevance of having a **homogeneous list of harmonised indicators** with a clear criteria.



- ✓ **The importance of information in electronic formats** for its massive and automated treatment.

THANKS FOR YOUR ATTENTION



APPENDIX



0. INVENTORY OF INDICATORS AND SOURCES

Selection of ESG indicators for the database – some examples



Nº de indicadores	124
E-Environmental	62
S-Social	44
G-Governance	18

1- Currently collecting (*)	5
2- Would like to collect	34
3- Not interest in its search	85
TOTAL	124

Indicator	Type (E,S,G)	Subtype	Metric	Standards applied	Informa	Regulation	Search on non-financial statements			Statistic Department	Level of interest
							Cons.	Indiv.	Level of difficulty		
Energy consumption within the org.	E	Energy	KWh, MWh, GWh, TWh, kJ, MJ, GJ, TJ, PJ	GRI 302-1 AECA	NO		YES	n.d.	1-Easy	1-Currently collecting (*)	1-High
Water consumption	E	Water	m3, Hm3, megalitros	GRI 303-5 AECA	YES		YES	n.d.	1-Easy	1-Currently collecting (*)	1-High
Direct GHG emissions (scope 1)	E	Green House Gases (GHG)	MtCO2e, TnCO2e	GRI 305-1 AECA	YES	NFRD	YES	n.d.	1-Easy	1-Currently collecting (*)	1-High
Hazardous waste	E	Waste	Tn	Old GRI 306-2	YES		YES	n.d.	1-Easy	2- Would like to collect	1-High
% renewable energy among the total	E	Energy	%	GRI 302-1	NO	NFRD	YES	n.d.	1-Easy	2- Would like to collect	1-High
Number of layoffs	S	Diversity, equality and well-being of staff	Nº	NO	YES		YES	n.d.	1-Easy	2- Would like to collect	1-High
Complaint channel	G	Corruption and bribery	Si/No/ND	NO	YES		YES	n.d.	1-Easy	2- Would like to collect	1-High

(*) Information available since 2019 for 30 Spanish Groups listed

❑ A web application with (semi-automatic) full-text search and storage capabilities has been developed.

- 1. The user selects a **year, company and indicator.**
- 2. The tool **searches the ontology** in company documents and presents an ordered list of paragraphs in order of relevance.
- 3. The user **validates the search results and stores the indicator value** into the database.
- 4. **Context information** (user, data, page and paragraph, terms and search results...) is saved, which will serve to **improve the automation** in the search process.

Restablecer términos de búsqueda

Términos de búsqueda:

agua consumo m3 303-5 fuentes reutilizada dato hídrico uso total subterránea m³ captación toneladas litros mar volumen hm río dulce marina lago millones caudal gast

514 resultados de búsqueda

Número de resultados:

10

Selección un documento:

6848836_MEMCON (100%)

Selección una página:

Todas

Selección un resultado:

1 (16 %)

Eliminar resultados duplicados

1 (documento 6848836_MEMCON, página 231, bloque 1, score 16 %): Consumo de agua [303] Consumo de Agua (m3) 2020 2019 Extracción de agua 37.737.355 20.677.055 Vertido de agua 2.663.210 2.199.031 Agua vendida 3.168 0 Total Consumo de agua (m3) 35.070.977 18.478.024 Nota: no se ha considerado como vertido el volumen de agua usado en el cultivo de arroz Consumo de agua por zonas de Stress Hídrico (m3) 2020 2019

231 de 279

Consumo de agua [303-5]

Consumo de Agua (m3)	2020	2019
Extracción de agua	37.737.355	20.677.055
Vertido de agua	2.663.210	2.199.031
Agua vendida	3.168	0
Total Consumo de agua (m3)	35.070.977	18.478.024

Usuario: CRISTINA GONZÁLEZ TAMAYO, q31432

Selección un ejercicio:

2020

Selección una prioridad:

Alta

Selección una empresa:

EBRO FOODS SA

NIF empresa: A47412333, Código interno: 140268

Selección uno o varios documentos:

6848836_MEMCON

Indicador

Consumo de agua.

Empresa localizada en zona de estrés respecto al agua.

% energía renovable entre el total.

Consumo de energía dentro de la organización.

Reducción del consumo energético

Emisiones directas de GEI (alcance 1).

Emisiones directas de GEI al generar energía (alcance 2).

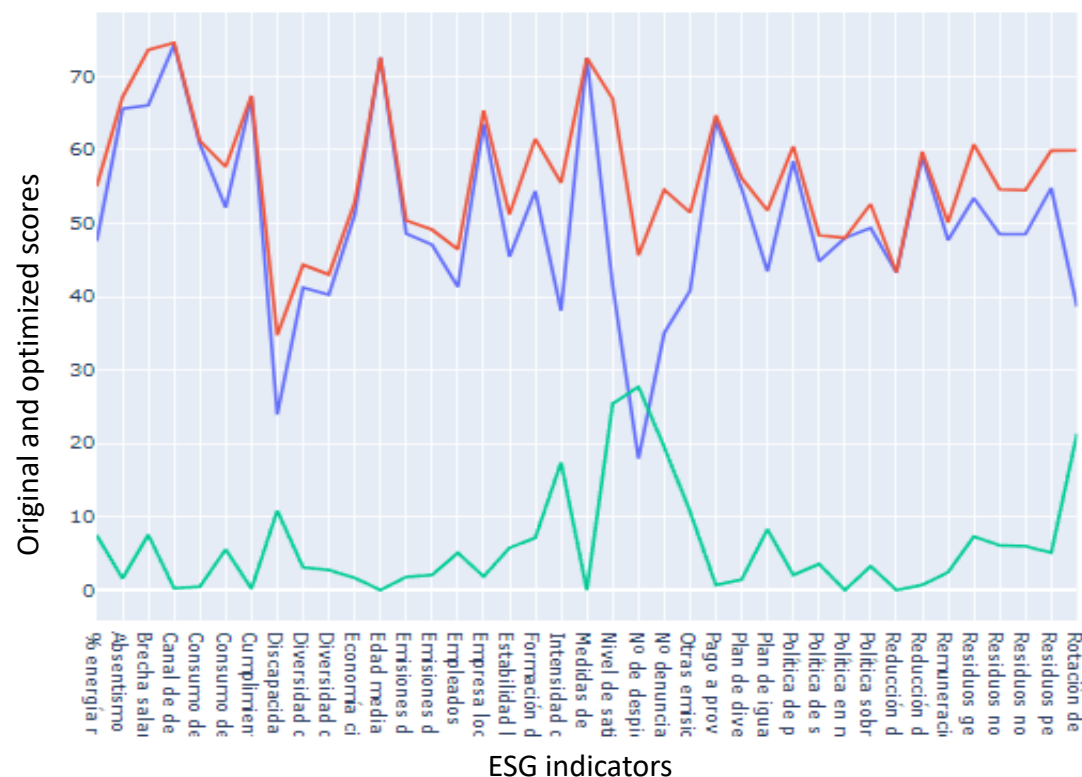
DIRECTORATE GENERAL ECONOMICS, STATISTICS AND RESEARCH

15

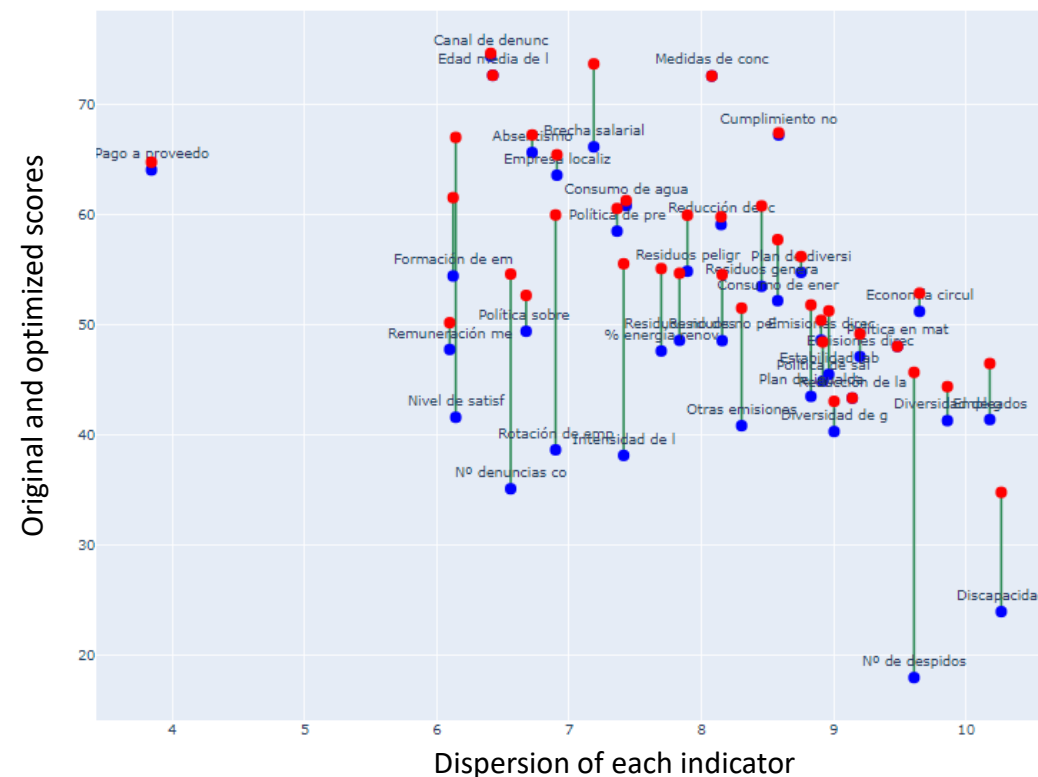
❑ Example of comparison and optimization of search queries:

- Original query: '*consumo de agua dulce salada subterránea m3*' -> score = **50%**
- New optimized query: '*consumo agua m3*', -> score = **60%**

- ❑ Scores of **original queries** (blue), **optimized queries** (red) and achieved **improvements** (green).



- ❑ Indicators with **lower dispersion** (more separable) achieve **higher scores**, as expected.



1

Understanding the behaviour of ESG indicators:

- Logical values (limit and comprehend the possible values of the indicators)
- Analysing the metrics in which indicators are reported (KWH, MWH, GJ...)
- Searching for alternative methods to obtain the data
- Value changes between exercises

2

Improving in the automatic search process:

- Suggestions of new words
- Removal of old words that could distort the search
- Sharing improvement in queries with Data Scientists

3

Increase quality of the database:

- Error corrections
- Delete inconsistent values
- Review correct metric consignment

0. PRELIMINARY RESULTS

Results and conclusions of the first extraction

- 20 employees were working part-time during 2 months to make the first

data ingestion:

- 10,000 records
- 39 indicators
- 164 groups
- 470 documents

- Visualization with PowerBI

