

# 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

11<sup>TH</sup> 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

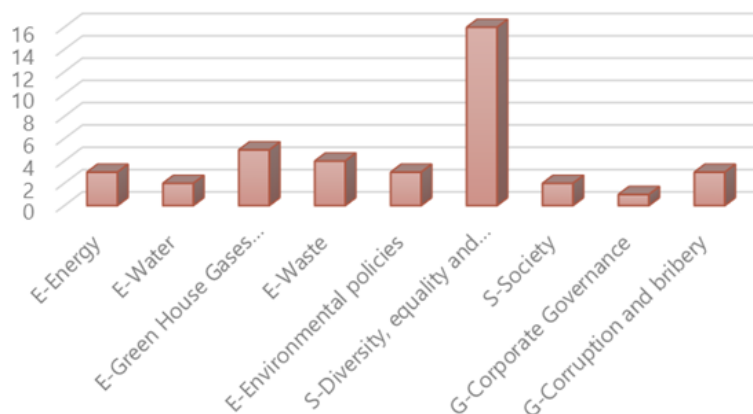
### 4. Conducting a practical research exercise involving six listed companies



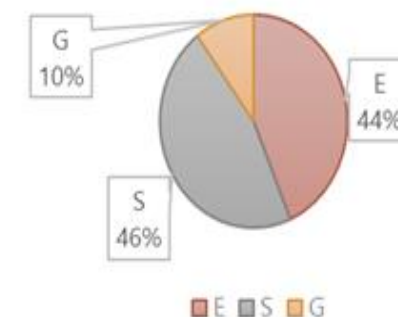
### 5. List of 39 ESG indicators to search for

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

Distribution of indicators by subtype

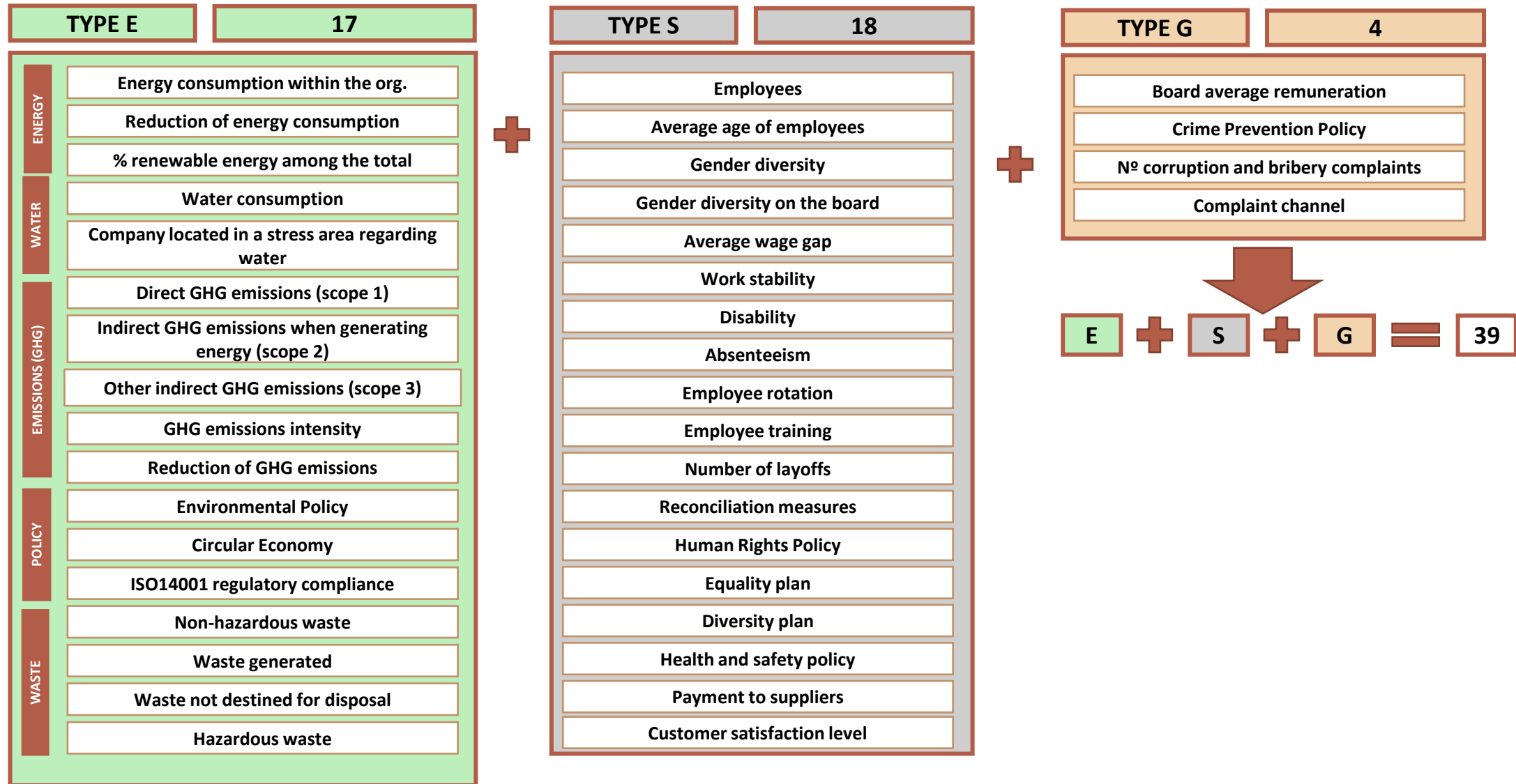


Distribution of indicators by type (ESG)



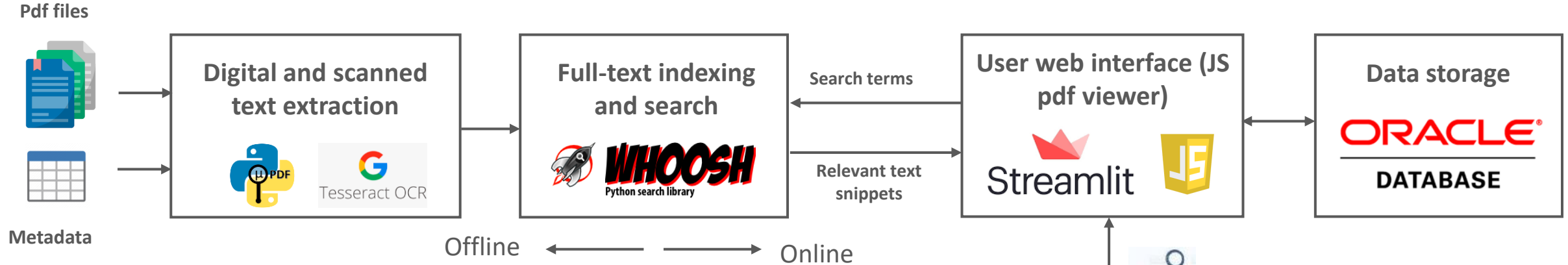
## 2. RESEARCHING AND ESTABLISHING THE RELEVANT INDICATORS

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 is 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, with a total consumption of 35,070,977 m3 in 2020 and 18,478,024 m3 in 2019.

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
<b>Total Consumo de agua (m3)</b>	<b>35.070.977</b>	<b>18.478.024</b>

(\*) 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



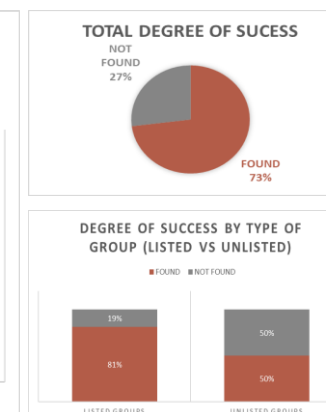
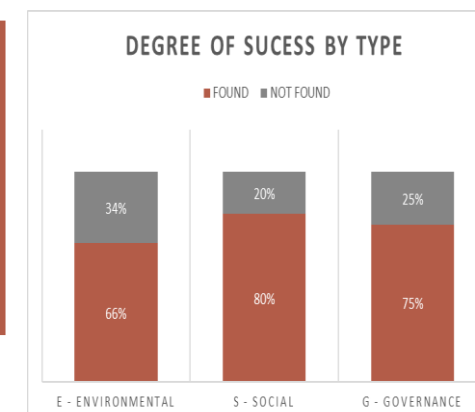
INTERNAL ENERGY CONSUMPTION BY PRIMARY SOURCE (TJ)*			
Fuel type	2017	2018	2019
Total ENDESA consumption	684,142	615,336	507,614

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

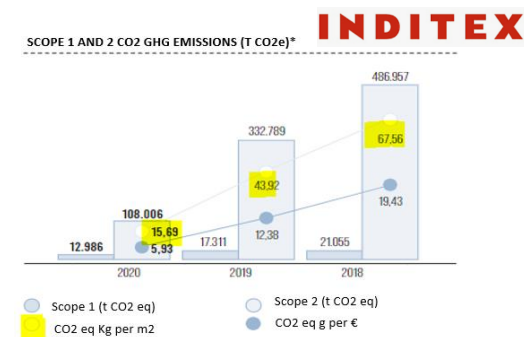
### LACK OF DIGITALISATION OF THE INFORMATION PRESENTED

- 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
<b>TOTAL</b>	<b>10,090,602</b>	<b>6,440,901</b>	<b>5,996,951</b>

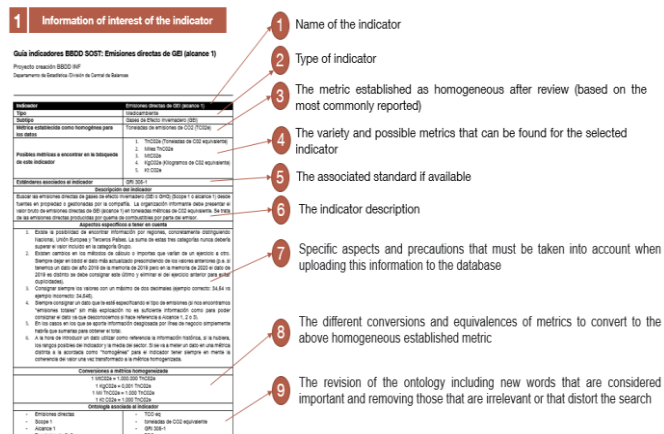
Fuel used in Stationary and Mobile sources (total) (GJ)

**euskaltel**

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

## Quality control of sustainability indicators

## 1 Information of interest on the indicator



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

**Analisis del indicador**

Descripción: describe tener en cuenta que el indicador no está siempre disponible para la misma empresa a lo largo de la serie histórica y, además, para determinados sectores puede no ser necesario o no tener alguna información. En cualquier caso, estos datos y resultados deben ser como fuente de apoyo y contrastar a la hora de medir el valor en la tabla y poder encontrar una idea de si el valor que vamos a introducir tiene alguna lógica no, hay que destacar que el valor se presentará para los años de "Historia pas. (último)" y con la entidad ya homogeneizada a "OCDE".

A continuación, se presenta la muestra por sectores utilizada para el análisis:

SECTOR	2000	2005	2010	2015	2017	2018
1. Agricultura	0	0	0	0	0	0
2. Minería y extracción	0	0	0	0	0	0
3. Industria manufacturera	0	0	0	0	0	0
4. Construcción	0	0	0	0	0	0
5. Comercio y hostelería	0	0	0	0	0	0
6. Servicios	0	0	0	0	0	0
7. Total	0	0	0	0	0	0

A continuación, se presenta la muestra media del indicador para la muestra analítica:

**Media (OCDE)**

SECTOR	2000	2005	2010	2015	2017	2018
1. Agricultura	0	0	0	0	0	0
2. Minería y extracción	0	0	0	0	0	0
3. Industria manufacturera	0	0	0	0	0	0
4. Construcción	0	0	0	0	0	0
5. Comercio y hostelería	0	0	0	0	0	0
6. Servicios	0	0	0	0	0	0
7. Total	0.000	0.000	0.000	0.000	0.000	0.000

Alcance: 1. Medio por año (OCDE)

1 A description of the type of information used and the current limitations so it can serve as a source of support and contrast when new indicators are loaded

2 The distribution of the sample that has been used by years and by sectors (in the case of the numerical ones)

3 The average values (in the homogenized metric) that will serve as a reference. Another table has been included with the maximum values to be able to establish normal ranges for each indicator (in the case of the numerical ones)

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.

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 corporate website with a navigation menu on the left and a main content area. The navigation menu includes links to 'Sustainability', 'Investor Relations', 'Media', 'Contact', and 'About Us'. The main content area features a 'Sustainability' section with a 'Sustainability Report' link, a 'Sustainability Dashboard' link, and a 'Sustainability Information' link. Below these links, there is a table with columns for 'Sustainability', 'Investor Relations', 'Media', 'Contact', and 'About Us'. The table contains various links and information related to these areas.

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.

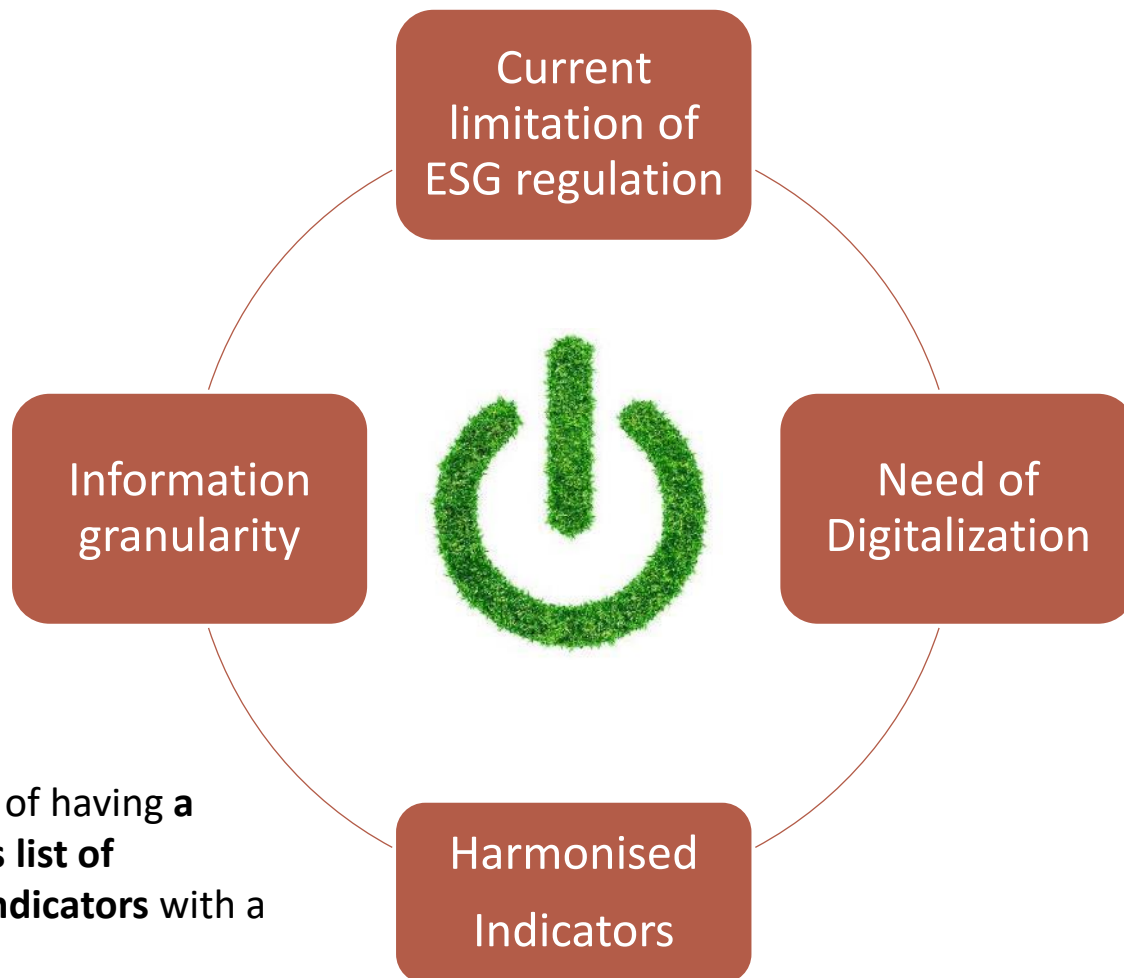
## 6. CONCLUSIONS

### 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

-

+

Seleccione un documento:

6848836\_MEMCON (100%)

Seleccione una página:

Todas

Seleccione 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

-

+

160%

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

Seleccione un ejercicio:

2020

Seleccione una prioridad:

Alta

Seleccione una empresa:

EBRO FOODS SA

NIF empresa: A47412333, Código interno: 140268

Seleccione uno o varios documentos:

6848836\_MEMCON

Indicador

0 Consumo de agua.

1 Empresa localizada en zona de estrés respecto al agua.

2 % energía renovable entre el total.

3 Consumo de energía dentro de la organización.

4 Reducción del consumo energético

5 Emisiones directas de GEI (alcance 1).

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

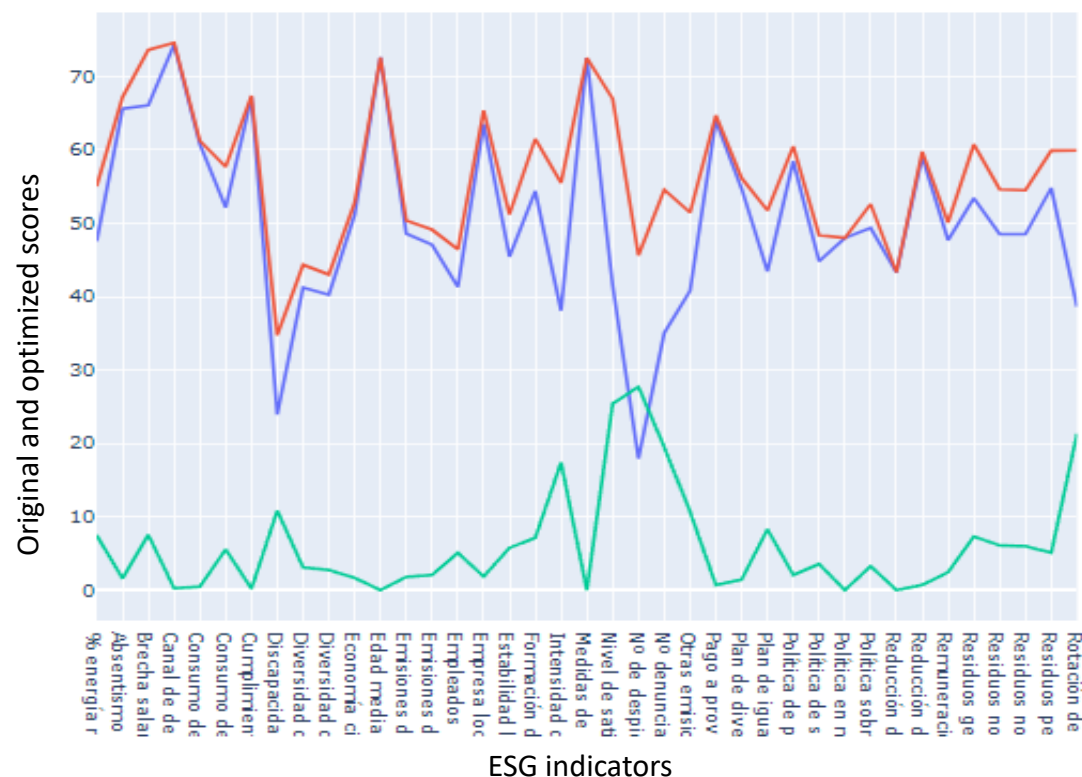
DIRECTORATE GENERAL ECONOMICS, STATISTICS AND RESEARCH

15

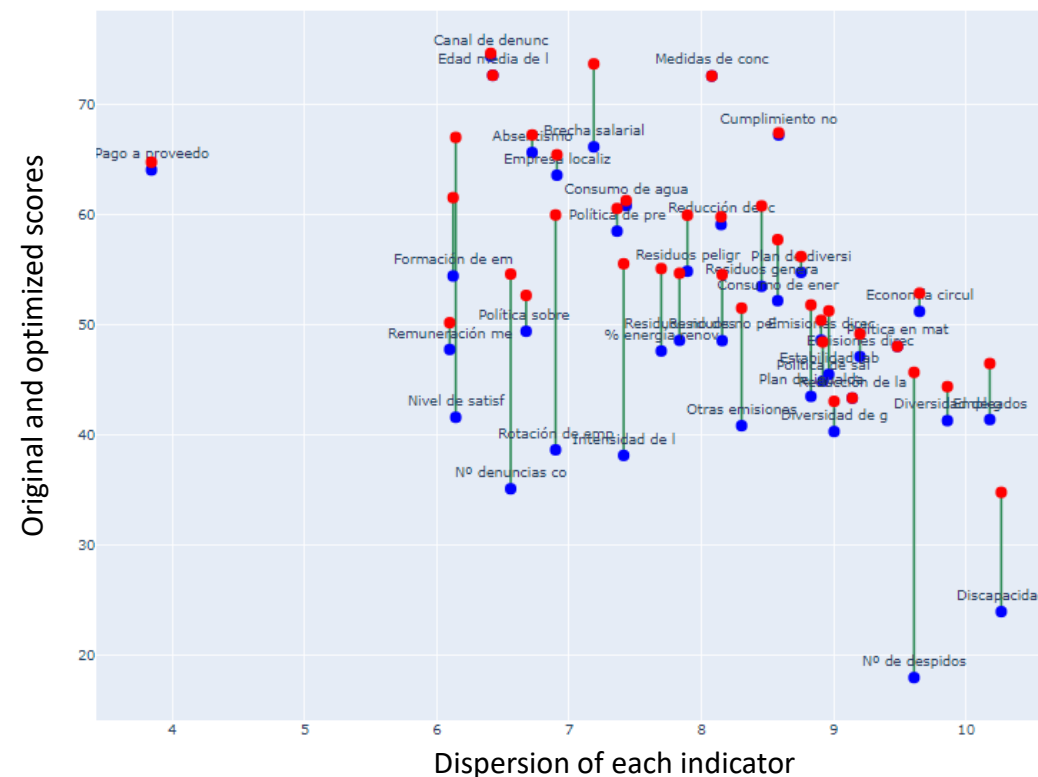
### ❑ Example of comparison and optimization of search queries:

- 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

