Typology of Data Quality Problems in the Corporate Reporting of GHG Emissions

A report by **[Green Data, Indicators, Algorithms (Green DIA)]**, funded by the Bavarian Research Institute for Digital Transformation (bidt) and **[Klimatkollen]**.

Abstract

This report presents a typology of data quality problems frequently found in corporate sustainability reports. The methodology links a conceptual approach (based on GHG Protocol principles) with an annotation framework for errors and quality issues in corporate sustainability reports. Our proposed typology of 30 errors and issues can assist annotators in the quality assessment of reports. In addition, the framework can be used for annotating data quality problems in automated text analysis with the help of Large Language Models (LLMs) or similar tools. Having performed the annotation, a quality statement can be assigned to a report that also groups the types and potential severity of the data quality problems. Furthermore, the quality assessment can aid in finding causes for low intercoder reliability and LLM accuracy.

Authors

Andreas Dimmelmeier (LMU Munich), Felicitas Sommer (Technical University Munich), Alexandra Palmquist (Klimatkollen)

Supplementary Materials
The full data quality problem typology as well as the annotation framework and annotation examples are available in the linked Open Science Foundation repository: https://osf.io/zwr26/

Feedback and further development:
The typology and the annotation framework outlined in this report are part of ongoing work at the GreenDIA consortium and at Klimatkollen to extract high quality data from corporate reporting. We highly welcome feedback or ideas for future development by researchers and practitioners.
If you want to get in contact please write to:

GreenDIA
Andreas Dimmelmeier, a.dimmelmeier@stat.uni-muenchen.de
Felicitas Sommer, felicitas.sommer@tum.de

Klimatkollen
Ola Spännar, ola@klimatkollen.se

Introduction

Natural Language Processing (NLP) and Large Language Models (LLMs) in particular have increasingly been deployed to process company reports to create structured databases of corporate Greenhouse Gas (GHG) emissions.

So far, LLM methods to extract emission values have – in an attempt to weed out hallucinations – been optimized for treating all reported GHG emissions values in corporate disclosures as the correct ones. This focus has, however, had the unintended side effect of treating emission disclosures of varying quality and credibility alike.[1] Moreover, by abstracting from contextual information about issues like the calculation methods, data sources, assumptions or consolidation approaches, data users are also deprived from finding the sources of errors and inconsistencies in aggregated databases that were created with the help of LLMs.

In this paper we want to bring the attention of analysts, as well as developers, and users of NLP methods to the data quality aspects of GHG emission extraction tasks. We do this by first introducing a typology of 30 emission reporting data quality problems that have been known to experts. These problems are often discussed in the literature, but so far have not been systematically brought together and categorized.

Additionally, we illustrate two use cases for our typology that can help human annotators (use case 1) as well as automated text processing applications (use case 2) to assess the data quality of company reports .

1. Use case 1: For human annotators we propose a checklist of questions corresponding to our 30 categories. Analysts can use this list to check which issues and errors are present in a report. We also see potentials of using these questions as prompts in automated LLM-based text processing. Moreover, a count of the number of present categories can be used as a vague indicator of the quality of a report.
2. Use case 2: Our second tool is a selection of text examples and images (e.g., screenshots of tables, graphs or other visuals) from company reports that we have annotated according to our typology. Again, these can be used by humans as examples for manual annotation or as inspiration to develop larger datasets to validate or fine-tune existing LLM-based extraction pipelines.

The paper first gives some background on data quality problems in corporate emissions reporting. Afterwards, we introduce our typology and discuss its links to existing frameworks like GHG principles as well as its conceptual background. Third, we illustrate both the checklist and the example annotations derived from our typology by applying it to real-world examples. We conclude by outlining applications as well as limitations and future research directions.

[1] Efforts by the NLP community of addressing data quality in reports through the detection of "greenwashing", while being related superficially, are going into a different direction as they are attempts to attribute a quality label to the claims made in a report rather than the disclosed data points.

Motivation: reporting errors and issues and their effects

Quality issues with companies' GHG disclosures are manifold and widely documented in the academic and industry literature. Among these issues there are trivial - but concerning - errors like the incorrect summation of GHG categories. Such calculation errors are surprisingly widespread prompting one study to question major corporations' "ability to use appropriate software or even basic tools like Excel spreadsheets to report their breakdowns accurately" (Garcia Vega et al. 2023: 4). On the other end of the spectrum of complexity, there are hard-to-identify reporting issues that are linked to the intricacies of disclosure rules. One example of these complicated issues are so-called "restatements", which denote the process of updating past emission values after a company's boundaries have changed (e.g., through M&A). As companies disclose different levels of detail on how they deal with issues like restatements (or none at all) even professional analysts from established data vendors in the financial industry struggle to identify the "correct" values (Kob and Dittrich 2024). The implications of such difficulties are also highlighted in the study by Busch et al. (2022: 357), which suggests that the extraction of different values (due to issues like restatements) could be an explanation for the inconsistencies found in the GHG databases from different commercial vendors.[2]

While restatements and other quality issues linked to the calculation methodologies or the interpretation of the GHG Protocol provisions (WBCSD and WRI 2011) arguably are related to limited capacities and experiences, researchers have also found instances where companies are purposefully disclosing incomplete or misleading information to cover up bad performance. Examples of such practices are arbitrarily limiting the disclosure of emissions to certain parts of the company or only disclosing emission intensities (e.g., CO_2 per turnover or passenger km) rather than absolute values (for examples of incomplete reporting see UNRISD 2023). Furthermore, with regards to Scope 3 emissions[3], companies might also only selectively report on certain sub-categories while excluding others. As demonstrated by one study focusing on technology companies, omissions of parts of the company (boundary incompleteness) and selective reporting of Scope 3 (activity incompleteness) can result in severe underreporting of emissions which would increase by 83% and 117% in the median respectively (Klaaßen and Stoll 2021).

Regardless of whether reporting quality problems happen because of inexperience or due to intentional choices of the reporting entity, from an annotator and data-user perspective, the most important question is how severe the problem is. Put differently, some errors or quality issues can lead to larger deviations from the true value (i.e., underreporting and in some cases overreporting) than others. Deducting the level of underreporting from types of errors or quality issues in sustainability reports is in general difficult and has not been done to our knowledge.[4] Summarizing

[2] Such inconsistencies are accentuated by the fact that the GHG Protocol, the leading framework for calculating and reporting emissions, offers considerable leeway for companies to choose their approach to emissions inventories, thereby impairing comparability, and is also plagued by conceptual inconsistencies (see Jia et al. 2022, Kasperzak et al. 2023 for a discussion).

[3] In the framework of the GHG Protocol, Scope 3 emissions are indirect greenhouse gas emissions that occur in a company's value chain. Scope 3 emissions include both upstream (i.e. suppliers) and downstream activities (i.e. product/service users) that are related but not directly controlled by the company.

[4] A notable exception in the field of NLP applications is the Sustainability Navigator, which also includes a list of standards' reporting provisions related to methods description and quality assurance in its 132 requirements. While this does not constitute an evaluation or sorting of the provisions, assessing whether and how a report complied with the provisions can give a user valuable information regarding data quality. https://sustainabilityreportingnavigator.com/#/requirements

and grouping errors and issues into different categories in a typology might thus be a first step to inform such calculations.

Why reporting quality matters for Climate NLP applications

Corporate sustainability reports contain unstandardized, unstructured text, requiring information extraction for assessment. ESG analysts at financial institutions and commercial data providers (like MSCI or Moody's), along with outsourced annotators, perform this extraction work (as documented in Kob and Dittrich's 2024 ethnographic study on calculation challenges). However, with the increasing availability of language processing tools at scale, both commercial providers and academics are relying more and more on automated text extraction.

While the varying quality of reports has not been lost on researchers in the field of "ClimateNLP", to date, research streams on quality assessment and the data extraction (e.g. extracting emission values) have remained separate. Accordingly, one line of inquiry focussed on quality addresses semantic and linguistic structures to detect "cheap talk" or "greenwashing" (Bingler at al. 2024, Hu et al. 2024, Gorovaia 2025, Ong et al. 2025). However, greenwashing identification has so far not been linked to quality principles from frameworks like GHG Protocol but is, instead, typically based on readability or claim-performance gaps. Data extraction is, meanwhile, the subject of other NLP pipelines (e.g., Webersinke 2021, Huang et al. 2023), which have improved in model accuracy, but researchers pay little regard to quality aspects regarding the original reports. The evaluation of these models is, in turn, sometimes based on emission data from commercial vendors. As these vendors use, however, also automated text processing to populate their databases, a problem of circularity arises as the model output is benchmarked against another (black-boxed) model output.

Studies thus focus either on extraction model accuracy or report quality, whereas from a data quality perspective a combination of both is needed. Ideally, the extracted numerical emission values should be accompanied by a confidence interval, which is based on the quality of the report and lets the user estimate the degree of underreporting. However, to date non-financial reports lack the standardization, scrutiny and disaggregation of information of financial reporting, thus making an immediate quality assessment by LLMs illusive. Instead, flagging these inconsistencies, errors, misinterpretable information and greenwashing (deliberately omitted or manipulated information) requires domain-specific expertise and experience. While human experts can trace these issues, it is not straightforward to simply add a module on data quality to existing extraction pipelines leveraging LLMs.

Moreover, taking LLM extraction errors (or low intercoder reliability from lay annotators) as a proxy for low report quality without further refinement might actually create misleading signals. This is because, as we experienced in different setups with LLM and human annotations (Beck et al. forthcoming, Klimatkollen 2024), low intercoder reliability and LLM errors often point towards quality problems of corporate reports - but can also stem from the opposite phenomenon, namely "better" reports with higher granularity and greater complexity such as like-for-like emissions disclosure in case of M&As.

In summary, we thus argue that the automated extraction of emission values needs to integrate domain-expertise based measures of report quality. Categorizing report quality would in this context be a first step towards finding different reasons for extraction failure. Furthermore, classifications on the confidence that one can have in extracted values as well as assessments of the possible extent of underreporting could start from such a typology of data quality problems.

Introducing the typology of 30 data quality problems

While the academic and industry literature has documented the sources of bad quality reporting such as ambiguous standards or corporate negligence as well as the potential effects including inconsistent databases and underestimated emissions, to our best knowledge,there is no comprehensive overview of reporting errors and issues.

To address this gap, we propose a list of 30 data quality problems (see table 1). We developed this list by reviewing critiques from the scholarly and grey literature (see Dimmelmeier et al. 2024: section 2 for a summary) as well as by aggregating and categorizing instances that we have come across by acting as annotators and validators in the context of LLM based data extraction tasks (see Beck et al. forthcoming, Klimatkollen 2024).

ID	Data Quality Problem Description	Issue or Error
E-1	Only reporting scope 1 and scope 2 (no scope 3 data)	Error
E-2	Only reporting total emissions	
E-3	Only reporting scope 1 and 2 together	
E-4	Not reporting scope 3 emissions per category (only total scope 3)	
E-5	Data in scope 3 is using other categories than those defined by the GHG protocol	
E-6	Not labelling scope 1, 2 and 3 clearly (eg using direct/indirect emissions as definitions)	
E-7	Reporting emissions by source, not mapped to categories	
E-8	No defined/stated base year	
E-9	Not disclosing method for scope 2 emissions (location based or market based)	
E-10	Calculation errors (eg not summarising categories correctly)	
E-11	Not recalculating base year or historical emissions in accordance with GHG Protocol guidelines	
E-12	Reporting biogenic emissions included in the scopes (should be reported separately)	
E-13	Different values in different parts of the report	
E-14	Only reporting intensity metrics (eg emissions per turnover, emissions per sold unit)	
E-15	Only presenting data in graphs and not with numbers	
I-1	Not reporting on all categories where there are significant emissions	Issue
I-2	Not indicating reason for omitting a category (e.g. no emissions, or no data available)	
I-3	Not indicating if a category is incompletely reported	
I-4	Not indicating if a value is estimated or calculated	
I-5	Not indicating limited assurance (report quality has been assessed by a third party)	
I-6	Not disclosing updated values from previous years when recalculations are made	
I-7	Operational boundaries not appropriately defined (eg excluding parts of the operations)	
I-8	Not reporting biogenic emissions	
I-9	Not disclosing emissions for all years from Base Year to Most Recent Year (technically not mandatory)	
I-10	Emissions reported in multiple locations (eg multiple reports, or different sections in the sustainability report)	
I-11	Not disclosing consolidation approach and rationale for it	

I-12	Disclosure of omitted assets or units	
I-13	Not disclosing all GHG types	
I-14	Voluntary disaggregation (e.g. reporting Scope 1-3 emissions by country, subsidiary or business segment)	
I-15	Visual and textual distance between emission numbers and technical documentation	

Table 1: Typology of data quality problems (reduced version)

Concept and background of the typology

Apart from describing the 30 data quality problems, we also sort them into five categories. These categories dichotomize between errors and issues (**Error or Issue**), assess annotation difficulty (**Annotation Difficulty**) and link the problems to the GHG Protocol (**Principles,** partly in **Broad Category**). In addition, they provide hints regarding the source of the problems (**Broad Category**). Finally and, perhaps most importantly, we introduce a more granular list that also allows for first assessments of the potential severity of the issue or error as well as its impact on extraction accuracy (**Fine Category**). Below we explain and illustrate the five categorizations. The full typology can be found on the OSF repository (Full_data_quality_problems_typology_1.0.csv).

Error or Issue

We divide the data quality problems into 15 errors and 15 issues (see table 1). Errors include factual errors such as summation errors and inconsistencies like the disclosure of different values for the same emission-year category on different pages. By far the greatest number of errors (12 out of 15), however, is related to non-compliance with the guidance of the GHG Protocol.[5] An example of such a non-compliance error is using other Scope 3 categories than those listed in the GHG Protocol (E-5 in Table 1)

Issues, meanwhile, pertain to problems like not providing sufficient information about calculation methodologies and estimations (I-4) or the approach taken to consolidate a company's boundaries (I-7). While not strictly an error in either the factual or the compliance sense, these issues, nevertheless, lead to interpretation difficulties, ambiguities or a lack of information necessary to judge whether a value is trustworthy.

Annotation Difficulty

The difficulty category sorts the data quality problems based on how difficult they are to spot for annotators. This category is largely based on qualitative evaluations of our experience with GHG annotation tasks and is split into the three levels of easy (13 observations), medium (12) and difficult (5).

[5] We are aware of the inconsistencies, issues (Jia et al. 2023, Tang et al. 2023) and politicized history (Walenta 2022) of the protocol, but nevertheless treat it as a reference for pragmatic reasons.

Principles

We also assign each problem to one of the 5 Principles of the GHG Protocol (WRI and WBCSD 2011: 7), which are summarized below.

- “Relevance” (3 observations): Ensure the GHG inventory appropriately reflects the GHG emissions (and removals, if applicable) of the company and serves the decision-making needs of users – both internal and external to the company.
- “Completeness” (6): Account for and report on all GHG emissions (and removals, if applicable) from sources, sinks, and activities within the inventory boundary. Disclose and justify any specific exclusions.
- “Consistency” (4): Use consistent methodologies to allow for meaningful performance tracking of emissions (and removals, if applicable) over time and between companies. Transparently document any changes to the data, inventory boundary, methods, or any other relevant factors in the time series.
- “Transparency” (10): Address all relevant issues in a factual and coherent manner, based on a clear audit trail. Disclose any relevant assumptions and make appropriate references to the accounting and calculation methodologies and data sources used.
- “Accuracy” (7): Ensure that the quantification of GHG emissions (and removals, if applicable) is systematically neither over nor under actual emissions (and removals, if applicable), and that uncertainties are reduced as far as practicable. Achieve sufficient accuracy to enable users to make decisions with reasonable assurance as to the integrity of the reported information.

However, not every error or issue could be clearly linked to one of the GHG Protocol principles. For example, reporting scope 1 and 2 summed up together instead of disclosing them separately could be categorized as relevance or transparency issue. We thus have come up with two additional frameworks for the clear categorization of errors or issues.

Broad Category

Our first additional category is a "Broad Category" that aggregates data quality problems into four categories.

- "Not following the GHG Protocol" (14 observations): The largest number of errors (as well as some issues) stem from non-compliance with the GHG Protocol. These include using different (sub)-categories than the "Scopes" from the Protocol as well as not stating and recalculating a base year in accordance with GHG Protocol provisions and applying consolidation approaches that do not comply with the protocol.
- "Misleading Reporting" (12): Issues (as well as some errors) that, while not strictly non-compliant with GHG Protocol, cast doubt on the reliability of the reported values. This category includes issues like the lack of information on the calculation methods, the specification of a company's operational boundaries and rationales for excluding certain assets or business units. In addition, errors related to confusing or even deceiving data presentation (explanatory endnotes far removed from tables, only graphics with no numbers) are sorted into this category.
- "Lack of quality control" (2): Factual errors that should have been weeded out if the report would have been reviewed thoroughly including calculation errors and inconsistent disclosures on different pages.
- "Reproducibility" (2): Relates to cases where the company provides sufficient detail and granularity that an independent party can judge the data quality or reproduce and trace back emissions calculations. While this is a "positive" issue, it can lead to extraction errors by LLMs.

Fine Category

The most fine-grained sub-categorization (Fine Category) comprises the centerpiece of our typology as the categories were developed to provide a first guidance regarding potential impacts of an error or issue on data extraction and reliability.

Fine Category Name (Number of data quality problems in category)	**Detailed explanation**	**Example data quality problem**	**Potential impact on extraction/ annotation accuracy**	**Potential impact on measurement accuracy (e.g. underreporting)**
Boundary specification (6)	Non-reporting on certain emission generating activities, business units, or subsidiaries as well as the non-provision of rationales why certain parts of a company are included or excluded. Disclosed values might hence not represent the whole company	No disclosure on whether joint ventures are included or excluded	**Low:** Contextual information on boundary specfication does not affect extraction of GHG emission values and units	**High:** Exclusions can affect large parts of a company's GHG emissions
Data aggregation (2)	Disaggregation of emission values by business units or countries (positive). Aggregation of emissions and other variables into intensity metrics (negative)	Only intensity metrics are disclosed (e.g. tCO2/units produced)	**High:** Multiple values in disaggregated disclosures might confuse models or annotators Intensity metrics might be wrongly annotated as absolute values.	**Low:** Disaggregation makes verification actually easier through potential linkage with other data
Calculation error (3)	The report contains different values for the same metric	Total Scope 3 is not equal to summation of Scope 3 sub-categories	**High:** Different numbers for the same metric can confuse annotators or models	**Low-High:** Errors might be very small or very large
Format (3)	Problems related to the presentation of information such as the structure of the report as well as problems with tables, images and footnotes	Emission disclosures for different scopes or other disaggregations are scattered throughout the report	**High:** Inconsistent or hard-to-interpret disclosures can lead to extraction failure	**Low:** Confusing report structures do not per se affect the quality of information
Description of method (4)	Description and updates of calculation methodologies, assumptions, data sources and assessments of data quality	No information on whether Scope 2 emission are market or location-based	**Low**: Contextual information on methods does not affect extraction of GHG emission values and units	**Medium-High:** Methodological specifications can have large effects
Missing (emissions) data (4)	No values for certain Scopes or years	Missing sub-categories in Scope 3	**Low:** Missing data results most likely in true negatives	**Medium**: Missing data can be an issue if reported categories are cherry picked or if trends cannot be assessed
Emissions categorization (8)	Emissions are disclosed and aggregated with categories that are different from the ones of the GHG Protocol	Only reporting Scope 1 and 2 emissions together	**Medium:** Depending on the annotation/extraction specifications either no values or "wrongly categorized" values will be extracted with the latter case leading to comparability issues across reports	**Medium:** Idiosyncratic categories can make comparisons and verification difficult

Table 2: Detailed description of "Fine Category"

Further details regarding the significance of the values of “Fine Category” for either report production or report analysis can be found in the figures and tables in Appendix of this report. There we also give an overview of how the five different categories we develop in the typology are interrelated.

Testing and use cases

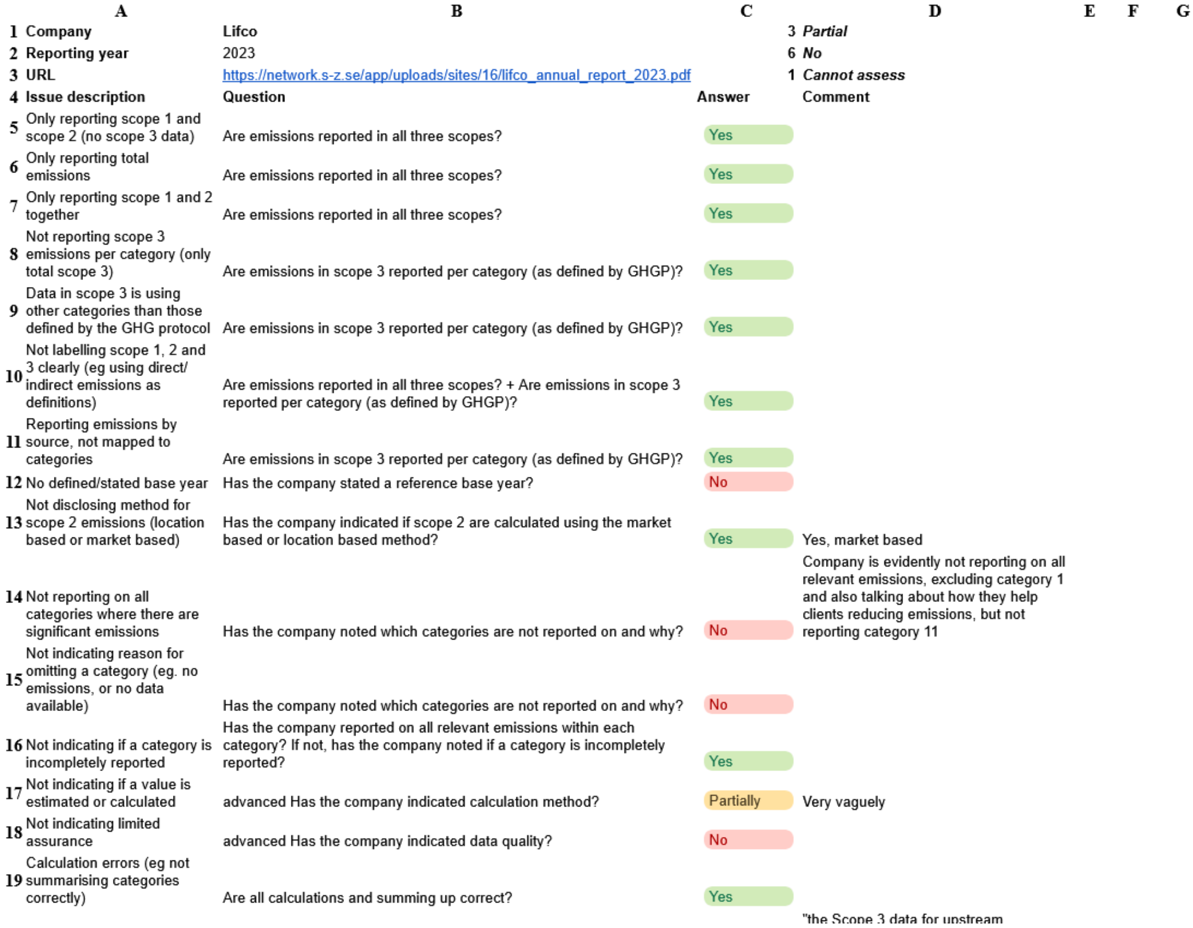
We contend that the typology can be applied to two different use cases. First, at the level of the report we can derive 30 questions from the issues and errors and record which and how many of them are present. Performing this analysis for a sample of company reports can deliver insights regarding the presence and variations of particular issues and errors across different time periods, sectors and geographies. Moreover, the overall reporting quality of any company, measured by the ratio of found categories divided by all 30 issues and errors, can be assessed.

Second, we can annotate a text chunk or screenshot as belonging to one of the 30 data quality problems. Such labelled data can then be used for validating error detection of multimodal AI models or for training human annotators.

To illustrate both use cases we tested the report-level and data quality problem-level annotation on a total of six company reports. We sampled reports that in previous analyses (Klimatkollen 2024, Beck et al. forthcoming) had been identified as particularly challenging by annotators (based on long annotation times and annotator comments). In this way we try to ensure that the illustrative cases show a comparatively high amount of data quality problems that can be labelled. Below we show an example of each annotation task. The report annotation framework (Annotation_framework_1.0.xlsx) as well as annotations of text chunks (text_chunk_annotation_examples.1.0.csv) and screenshots (folder: annotated screenshots) can be found on the OSF repository.

Report annotation tool example: SME manufacturing companies’ holding (Europe, 2023)

The spreadsheet tool prompts the annotator to answer each of the 30 questions with “Yes”, “No”, “Partially” or “Cannot assess”. These questions are worded in a way that “Yes” values indicate the absence of errors and issues or the presence of “positive issues”, whereas “No” values represent the opposite (see figure 2). In addition, annotators get a 31st row where they can indicate whether the report displays any additional quality problems and comment in case this question is answered negatively. Annotators can also elaborate a rationale for their choice in a comment section for the other 30 questions. We do not provide annotators with specific instructions or search strategies regarding the answering of the questions. This is because the targeted audience of the tool are “expert annotators” that have some familiarity with GHG accounting and the reading of reports but so far lack a way to structure their observations about data quality. (Lay) annotators without such a background knowledge are directed to resources like our annotation tutorial (Beck et al. forthcoming) to familiarize themselves with the extraction of GHG emission values.



	A	B	C	D	E	F	G
1	Company	Lifco	3	*Partial*			
2	Reporting year	2023	6	*No*			
3	URL	https://network.s-z.se/app/uploads/sites/16/lifco_annual_report_2023.pdf	1	*Cannot assess*			
4	Issue description	Question	Answer	Comment			
5	Only reporting scope 1 and scope 2 (no scope 3 data)	Are emissions reported in all three scopes?	Yes				
6	Only reporting total emissions	Are emissions reported in all three scopes?	Yes				
7	Only reporting scope 1 and 2 together	Are emissions reported in all three scopes?	Yes				
8	Not reporting scope 3 emissions per category (only total scope 3)	Are emissions in scope 3 reported per category (as defined by GHGP)?	Yes				
9	Data in scope 3 is using other categories than those defined by the GHG protocol	Are emissions in scope 3 reported per category (as defined by GHGP)?	Yes				
10	Not labelling scope 1, 2 and 3 clearly (eg using direct/ indirect emissions as definitions)	Are emissions reported in all three scopes? + Are emissions in scope 3 reported per category (as defined by GHGP)?	Yes				
11	Reporting emissions by source, not mapped to categories	Are emissions in scope 3 reported per category (as defined by GHGP)?	Yes				
12	No defined/stated base year	Has the company stated a reference base year?	No				
13	Not disclosing method for scope 2 emissions (location based or market based)	Has the company indicated if scope 2 are calculated using the market based or location based method?	Yes	Yes, market based			
14	Not reporting on all categories where there are significant emissions	Has the company noted which categories are not reported on and why?	No	Company is evidently not reporting on all relevant emissions, excluding category 1 and also talking about how they help clients reducing emissions, but not reporting category 11			
15	Not indicating reason for omitting a category (eg. no emissions, or no data available)	Has the company noted which categories are not reported on and why?	No				
16	Not indicating if a category is incompletely reported	Has the company reported on all relevant emissions within each category? If not, has the company noted if a category is incompletely reported?	Yes				
17	Not indicating if a value is estimated or calculated	advanced Has the company indicated calculation method?	Partially	Very vaguely			
18	Not indicating limited assurance	advanced Has the company indicated data quality?	No				
19	Calculation errors (eg not summarising categories correctly)	Are all calculations and summing up correct?	Yes				
				"the Scope 3 data for upstream			

Figure 1: Screenshot: Filled out question list for annotators

After finishing the annotation, the tool summarizes the counts of data quality problems on a results page (figure 3). In our example 21 out of the 31 categories are answered with a “Yes” indicating that the annotator has not found any corresponding issues or errors for those questions. Meanwhile, the 10 questions where the annotator found the presence of issues or errors are listed on the results sheet alongside comments and an explanatory text for the respective categories. In our example, the annotator has noted that recalculations of emissions do not follow GHG Protocol guidance (row 14) and quoted a text chunk indicating the inconsistencies in the comments section. Moreover, the annotator points out that consolidation approach is not disclosed transparently while noting that it “looks like operational control” (row 18).

NAME OF TAB -->	**Assessment - Lifco 23**	
Company	Lifco	
Reporting year	2023	
URL	https://network.s-z.se/app/uploads/sites/16/lifco_annual_report_2023.pdf	
Score	21 out of 31 issues avoided!	
Potential improvements	10 areas to improve :)	
Identified issues	**Comments**	**Why is this important?**
No defined/stated base year		Stating a base year is crucial to know from which year the company has comparable data. Otherwise, earlier reports may be seen as comparable even if they are not.
Not reporting on all categories where there are significant emissions	Company is evidently not reporting on all relevant emissions, excluding category 1 and also talking about how they help clients reducing emissions, but not reporting category 11	This is crucial to understand if the company has done a complete calculation of their emissions, or if they have done a partial calculation for some reason.
Not indicating reason for omitting a category (eg. no emissions, or no data available)		This is very important as it notes the difference between categories where the company has no emissions, and categories where the company has not been able to calculate their emissions, even if they may be significant.
Not indicating if a value is estimated or calculated	Very vaguely	It is very helpful if the company discloses the calcuation method, for example, using spend based data to estimate the footprint or if data was compiled from suppliers. This indicates data quality, and can serve as a reference for other companies, moving the industry forward in improving calculations and harmonising methods
Not indicating limited assurance		If data quality is indicated, the reader can better understand what is indicative and what is closer to the truth. This can help understand where the calculations are likely to be impoved and adjusted in the future, and what is more robust.
Not recalculating base year or historical emissions in accordance with GHGP guidelines	"the Scope 3 data for upstream transportation and biogenic emissions in 2022 was found to contain inaccuracies and have been omitted from the 2023 report. In 2023, Scope 3 emissions from employee commuting were included. Scope 3 data for the years 2019–2022 was also adjusted to take account of employee commuting." So some is excluded which doesn't seem like a great idea, but also cat7 is retroactively adjusted	If your company, or calculation method, has undergone a change that signifcantly impacts the previously reported number, it is important to recalculate the base year. This is so that emissions can be fairly compared from one year to the next - it may otherwise look like they increased, if the calculations become more inclusive, ot unfairly look like emissions were reduced when parts of the operations are divested.
Not disclosing updated values from previous years when recalculations are made		If historical emissions have been recalculated, it is appreciated if this is disclosed. The company will showcase their transparency by making note of what has changed, what the previous value was, and what the restated emissions are.
Operational boundaries not appropriately defined (eg excluding parts of the operations)		For complex organisations, stating the operational boundaries ensures an understanding of what is included, what is excluded, and why. This can be relevant for example if the company has operations in different countries and markets, or has shared ownership of some parts of the company.
Not disclosing emissions for all years from BY to MRY (technically not mandatory)	Yes but not scope 3 per category, and also there is no base year...	By presenting all emissions from the base year to the most recent year, the company allows for an understanding of the full volume of reported emissons. If emissions are consistently presented in this way, the reader does not have to rely on old reports with data that may be outdated to get the full overview of a company's emissions. This is important to be able to understand year on year changes, and not just the average trend.
Not disclosing consolidation approach and rationale for it	Looks like operational control to me (group level)	

Figure 2: Screenshot: Results page of the online annotation tool

Examples: Annotated texts and screenshots

While the report-level annotation tool is directed at the analysis of companies, the data quality problem-level annotation provides more detailed insights into how the respective issue or error manifests itself and can be detected. As reports are multimodal documents making use of text, graphics, and tables, we provide example annotations for both text and image (screenshot) data formats.

In table 3, we present two entries from the file with text annotations (see repository) to showcase how text chunks can be labelled. The first text chunk pertains to the data quality problem of *Only reporting intensity metrics* (E-14), which in the example is related to a company only disclosing tonnes of CO2 divided by million pounds sterling. The second example is about *Not disclosing updated values when recalculations are made* (I-6) and features a text chunk explaining that due to updates in emission factors and calculation methods “it is not possible to compare greenhouse gas emissions between 2023 and 2022”.

annotated_text	data_quality_problem	company_report_name	report_url
The Group monitors its greenhouse gas emissions and uses **CO2 emissions per £ million of turnover as a measure of its environmental intensity**. In the year ended 31 March 2020, total energy use from electricity, gas and fuel was 9.42 TWh 1, equating to 2.5 kWh per £ turnover. The associated carbon emissions from this energy use were 2.3 million tonnes 1, equating to 622 tonnes per £ million turnover.	E-14	dart_group_2020	https://www.jet2plc.com/-/media/Jet2/Flights/pdf/Jet2_PLC/Annual_report_2020
AddLife's greenhouse gas emissions increased between 2022 and 2023. The main reason is better data coverage; for example, the use of district heating is now included, whereas in previous years it was difficult to obtain that information from our landlords. The increase in emissions is also due to an increase in our energy consumption, especially fuel consumption from our fleet of vehicles, which travelled further in 2023 than in 2022, as well as a refrigerant leak at one of our subsidiaries that had a significant climate impact. We have also updated our emission factors and calculation methods.**Therefore, it is not possible to compare greenhouse gas emissions between 2023 and 2022.** Our climate calculations comply with the GHG Protocol Corporate Standard.	I-6	addlife_2023	https://www.add.life/media/nwepx4r4/addlife-annual-report-2023.pdf

Table 3: Annotated text excerpts

Regarding the annotation of images, figure 4 shows the screenshot of a table that we annotated as an instance of the data quality problem *Not reporting on all categories where there are significant emissions*. This labelling is motivated by the inconsistent treatment of Scope 3 emissions sub-categories with the annotator noticing that “[categories] 8, 10, 15 are marked as “0”, but 13 is blank (see image)”.

Greenhouse gas inventory: scope emissions 1, 2 & 3

Sum of GHG emissions (tonnes CO_2 eq)	FY16	FY17	FY18	FY19	FY20	FY21	FY22	FY23
Scope 1 On-site generation, fuel combustion and refrigerants	122,975	115,729	110,893	82,962	70,978	68,714	78,371	61,467
Scope 2								
Purchased electricity & heating								
Location-based	430,830	378,183	413,843	405,406	377,651	369,515	371,606	283,571
Market-based	278,014	220,216	269,193	158,879	92,860	76,987	28,804	17,621
Scope 3								
1. Purchased goods and services	15,981,148	16,197,335	17,209,516	17,004,122	15,264,259	16,562,384	16,397,911	13,987,358
Food ingredients	963,022	899,309	974,772	891,476	721,033	700,534	867,276	861,316
Materials	12,104,912	12,292,274	13,238,396	13,087,259	11,835,942	13,639,662	13,477,895	11,215,033
Production	2,856,825	2,938,986	2,947,516	2,968,840	2,671,931	2,177,642	2,007,006	2,183,274
Retail equipment & co-worker clothing	56,389	66,766	48,832	56,547	35,354	44,546	45,733	43,358
2. Capital goods	199,633	193,558	348,943	316,464	188,116	126,622	114,448	286,463
3. Fuel- and energy-related activities	108,960	71,947	97,880	60,845	55,585	49,334	41,527	31,325
4. Upstream transportation and distribution	1,253,607	1,387,283	1,463,193	1,385,675	1,188,797	1,358,230	1,260,073	1,006,777
5. Waste generated in operations	53,190	79,700	103,207	87,552	51,432	46,654	36,179	34,069
6. Business travel	145,809	140,092	148,713	113,583	54,336	13,457	28,818	52,848
7. Employee commuting	352,820	376,863	393,447	397,395	388,376	427,260	423,540	397,390
8. Upstream leased assets	0	0	0	0	0	0	0	0
9. Downstream transportation and distribution	2,132,296	2,175,531	2,263,626	2,263,156	1,975,340	1,976,729	1,886,727	1,811,155
Customer travel	1,940,864	1,956,449	2,013,753	1,954,469	1,708,100	1,645,102	1,668,882	1,606,902
Home deliveries	191,432	219,082	249,873	308,687	267,240	331,627	217,845	204,253
10. Processing of sold products	0	0	0	0	0	0	0	0
11. Use of sold products	7,990,576	7,467,481	7,230,826	6,384,018	5,743,429	5,534,869	4,390,868	3,819,585
Appliances	1,780,111	1,756,626	1,777,768	1,752,255	1,529,837	1,653,852	1,309,509	1,056,961
Candles	71,946	68,051	66,232	58,792	53,269	56,031	51,614	41,309
Lighting	6,138,519	5,642,804	5,386,826	4,572,898	4,157,939	3,824,338	3,029,556	2,721,315
Home electronics	0	0	0	73	2,384	649	189	0
12. End-of-life treatment of sold products	1,686,746	1,668,840	1,783,251	1,782,493	1,635,301	1,771,085	1,619,666	1,480,157
13. Downstream leased assets								
14. Franchises	572,923	581,000	675,573	626,581	632,016	585,336	510,644	421,964
15. Investments	0	0	0	0	0	0	0	0
Grand total (For scope 2 emissions, the market-based value is used for purchased electricity and heating)	30,878,697	31,417,301	32,098,261	30,663,726	27,340,825	28,597,660	26,817,574	23,723,800
Outside the scopes Biogenic emissions (from on-site fuel combustion)	430,126	363,543	428,057	485,710	452,485	443,576	495,822	370,009

Figure 3: Annotated Table: Not reporting on all categories where there are significant emissions

In the image folder on the repository we provide 13 screenshots that we annotated according to the data quality problems found therein to illustrate how the typology might be used to train humans and automatized extraction pipelines to detect them.

Conclusions, limitations and future research

In this technical report we highlighted that human annotators and NLP applications need to pay attention to data quality and not only extraction accuracy. As regulatory efforts to increase data quality on the report producers side through greater standardization and auditing such as the the EU’s Corporate Sustainability Reporting Directive (CSRD) are currently in an uncertain political state, it remains all the more important for data processors and users to develop approaches for judging the reliability of disclosed information.

To take a first step to the systematic assessment of data quality problems in reporting we propose a typology of 30 errors and issues and outline two possible use cases for human annotators and automated data extraction. Currently, both the typology and the resources in the repository are work-in-progress. We thus invite academics as well as practitioners to test, review and expand the typology as well as the spreadsheet annotation framework and the text and image annotations.

References

Beck, J. et al. (2025). *Gold Standard and Annotation Dataset for CO2 Emissions Annotation* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.14833375

Bingler, J.A. *et al.* (2024) 'How cheap talk in climate disclosures relates to climate initiatives, corporate emissions, and reputation risk', *Journal of Banking & Finance*, 164, p. 107191. Available at: https://doi.org/10.1016/j.jbankfin.2024.107191.

Busch, T., Johnson, M., & Pioch, T. (2022). Corporate carbon performance data: Quo vadis? *Journal of Industrial Ecology*, *26*(1), 350–363. https://doi.org/10.1111/jiec.13008

Dimmelmeier, A.; Sommer, F., & Cooke, D. (2024). Reframing the EU landscape for corporate sustainability information. *GreenDIA* and *Sustainable Finance Observatory (formerly 2° investing initiative).* https://sustainablefinanceobservatory.org/resource/reframing-the-eu-landscape-for-corporate-sustainability-information/

Klimatkollen 2024. *Corporate Climate Checker – An analysis of 150 major Swedish companies' climate reporting 2023*. https://klimatkollen.se/reports/2024-08_CorporateClimateChecker.pdf

García Vega, S., Hoepner, A. G. F., Rogelj, J., & Schiemann, F. (2023). *Abominable Greenhouse Gas Bookkeeping Casts Serious Doubts on Climate Intentions of Oil and Gas Companies* (SSRN Scholarly Paper 4451926). https://doi.org/10.2139/ssrn.4451926

Gorovaia, N. and Makrominas, M. (2025) 'Identifying greenwashing in corporate-social responsibility reports using natural-language processing', *European Financial Management*, 31(1), pp. 427–462. Available at: https://doi.org/10.1111/eufm.12509.

Hu, P. *et al.* (2024) 'Peeking into Corporate Greenwashing through the Readability of ESG Disclosures', *Sustainability*, 16(6), p. 2571. Available at: https://doi.org/10.3390/su16062571.

Huang, A.H., Wang, H. and Yang, Y. (2023) 'FinBERT : A Large Language Model for Extracting Information from Financial Text*', *Contemporary Accounting Research*, 40(2), pp. 806–841. Available at: https://doi.org/10.1111/1911-3846.12832.

Jia, J., Ranger, N., & Chaudhury, A. (2022). *Designing For Comparability: A Foundational Principle of Analysis Missing In Carbon Reporting Systems* (SSRN Scholarly Paper 4258460). https://doi.org/10.2139/ssrn.4258460

Kasperzak, R., Kureljusic, M., Reisch, L., & Thies, S. (2023). Accounting for Carbon Emissions—Current State of Sustainability Reporting Practice under the GHG Protocol. *Sustainability*, *15*(2), 994. https://doi.org/10.3390/su15020994

Klaaßen, L., & Stoll, C. (2021). Harmonizing corporate carbon footprints. *Nature Communications*, *12*(1), 6149. https://doi.org/10.1038/s41467-021-26349-x

Kob J & Dittrich J. (2024). Carbonising financial assets: between commodification and assetization of carbon relations. Working paper: https://warwick.ac.uk/fac/soc/wbs/research/ikon/research/climate-finance/publications-and-working-papers/kob__dittrich_2024_carbonising_financial_assets_wcfr_working_paper.pdf

Tang, K., Shrimali, G., & Christiaen, C. (2023). *Data Quality Considerations for Estimating Financed Emissions*.

Walenta, J. (2021). The making of the corporate carbon footprint: The politics behind emission scoping. *Journal of Cultural Economy*, *14*(5), 533–548. https://doi.org/10.1080/17530350.2021.1935297

WRI, WBCSD (2011). The greenhouse gas protocol: A corporate reporting and accounting standard - revised version.

Ong, K. *et al.* (2025) ‘Explainable natural language processing for corporate sustainability analysis’, *Information Fusion*, 115, p. 102726. Available at: https://doi.org/10.1016/j.inffus.2024.102726.

UNRISD. (2023). Indicators That Matter Toward Authentic Sustainability Reporting. Research and Policy Brief 40. May 2023. https://cdn.unrisd.org/assets/library/briefs/pdf-files/2023/rpb40-sdpi-indicators-that-matter.pdf

Webersinke, N. *et al.* (2021) ‘ClimateBert: A Pretrained Language Model for Climate-Related Text’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2110.12010.

Appendix: Additional information on the typology

Table 4 shows at which step of the report creation (corporation) or extraction process (analysts) the issue or error appears and how it is translated into an issue of trust.

	Report creation			Analysis of report		Confidence (prelim.)
	Concept	Calculation	Representation	Extraction	Interpretation & Adoption	Trust
Operational Boundaries	x				x	low
Calculation Error		x			x	medium
Format			x	x		medium
Method Description			x		x	low
Data Aggregation		x			x	low
Missing Emissions Data		x			x	medium

Table 4: Further details on selected values from Fine Category

The Sankey plot displayed in figure 4 gives an idea about the relations between the different categories.

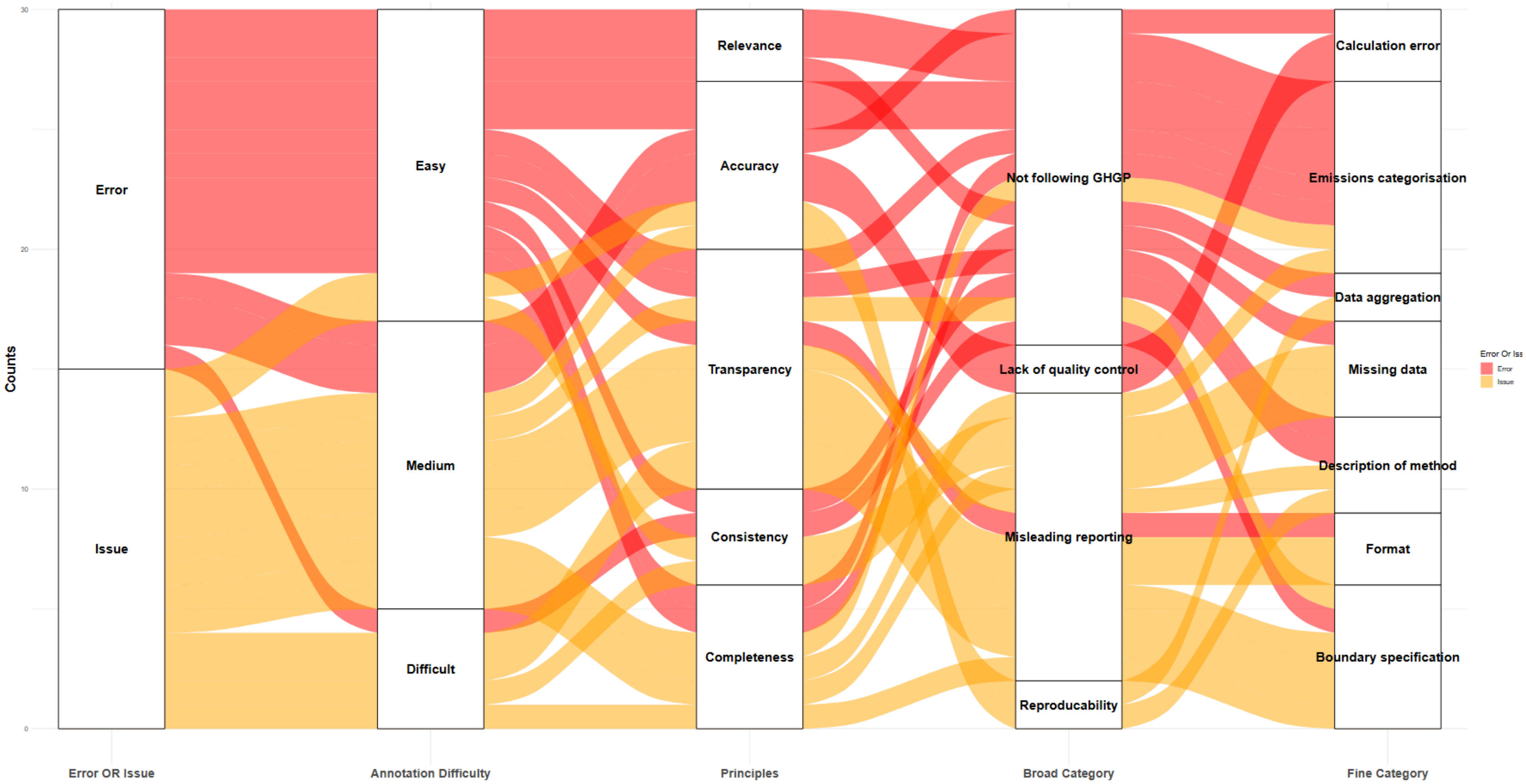


Figure 4: Relations between categories

The bar plot in figure 5 displays the relations between the well-known GHG principles and the “Fine Category” groupings proposed in this paper.

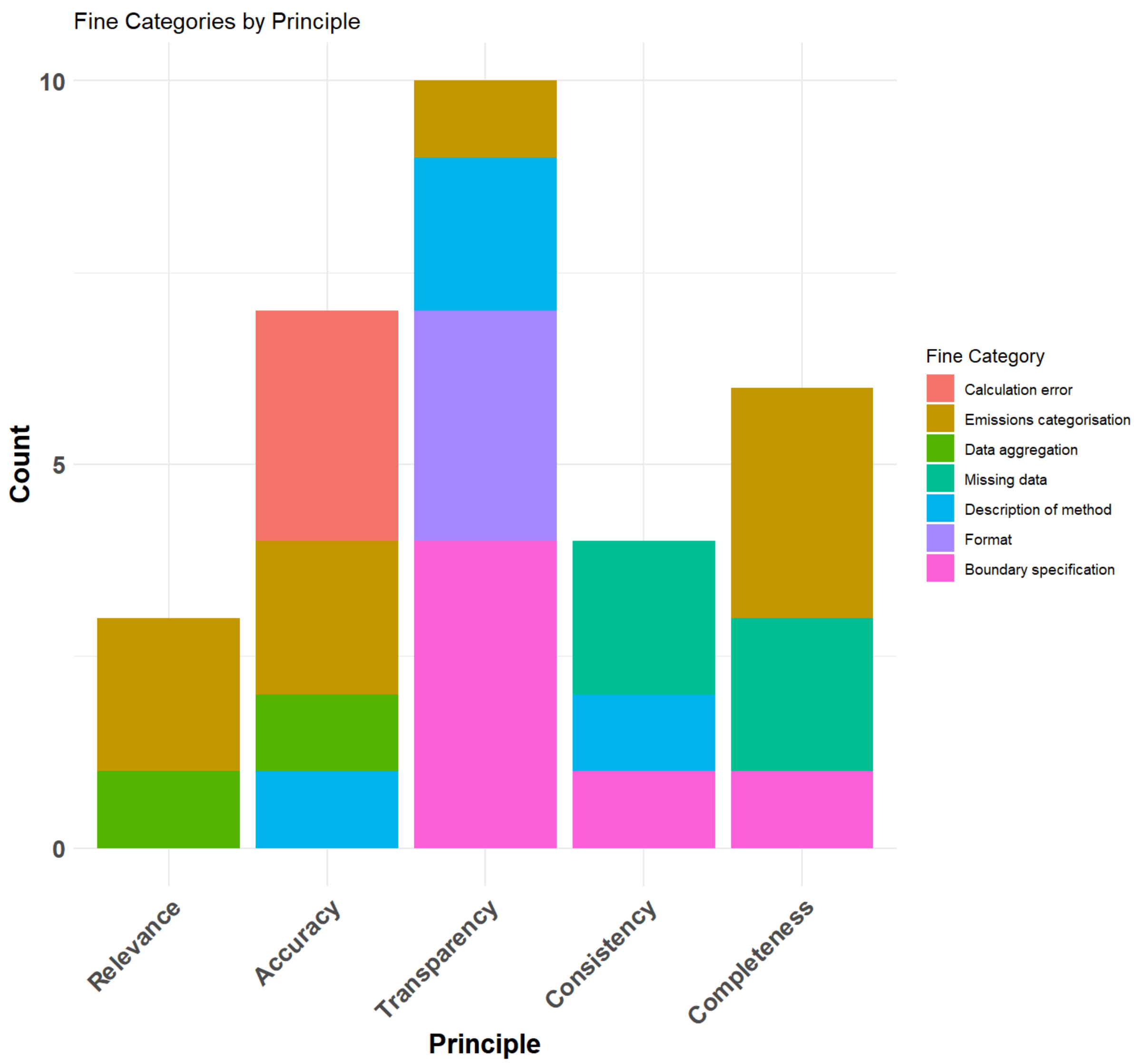


Figure 5: Relations GHG principles and Fine Category values