

PCF Input Flat Files - Validation Process

This page is dedicated to documenting the data validation processes for the PCF flat files data uploading, with the goal of improving data quality within our database. Through this effort, our aim is to cultivate more accurate and realistic results in PCF calculations.

1. Purpose

The purpose of this documentation is to ensure consistency and accuracy for the flat files uploading process. By adhering to these guidelines, we aim to maintain high data quality standards, which are crucial for reliable decision-making and operations.

2. Scope

This document covers the validation processes for the following flat files, which are used for PCF calculations.

Mapping Files

1. [PCF_InputFile_ef_supplier](#)
2. [PCF_InputFile_ef_rm_mapping](#)
3. [PCF_InputFile_sp_proxy](#)
4. [PCF_InputFile_Substitution_BOM_Imep](#)
5. [PCF_InputFile_ef_wastes_sites](#) (multiple sheets)
6. [PCF_InputFile_ef_transp_routes](#)
7. [PCF_InputFile_energy_mapping](#) (multiple sheets)
8. [PCF_InputFile_Energy_contribution_correction](#)
9. [PCF_InputFile_biogenic_masterdata](#) (multiple files)
10. [PCF_InputFile_MassBalance_substitution](#)
11. [PCF_biogenic_FP_manual](#)
12. [PCF_FP_manual](#)

Validation Files

1. [PCF_Validation](#)
2. [Biogenic Validation](#)

Ecoinvent Database Files

1. [GEO_GCP](#) (Geographical classification master table)
2. [AO_GCP](#) (Activities overview from the Database Overview xls file)
3. [LCIA_GCP](#) (Emissions factors by activity)

Configuration Files

1. [PCF_InputFile_Master_Unit_conversion](#)
2. [Advanced_GBU_Access_rights_PROD_](#)
3. [Config_default_PDS_DQR_PROD_](#)

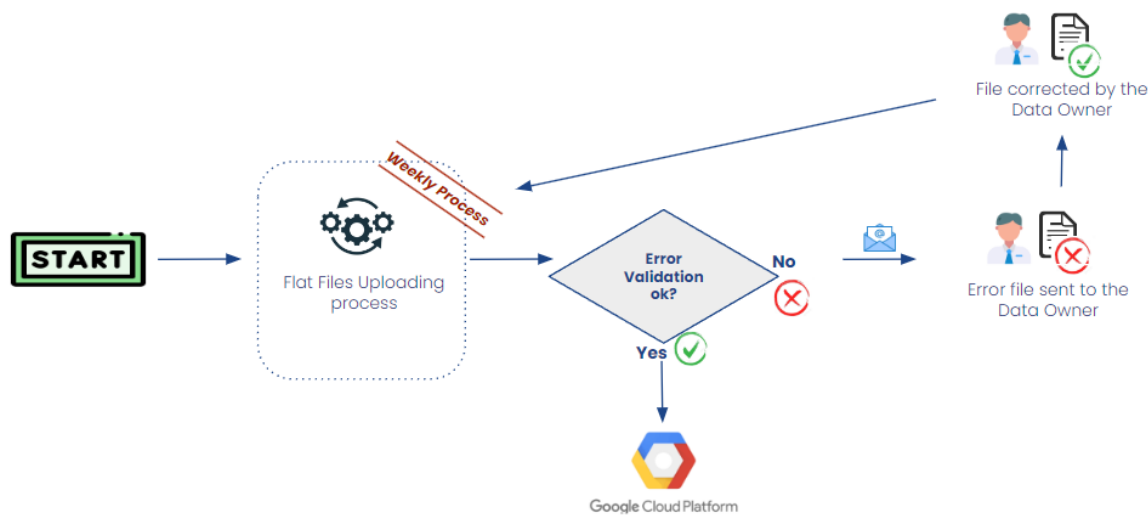
3. Importance of Data Quality

Data quality plays a pivotal role in the success of any data-driven organization. Poor data quality can lead to erroneous insights, inefficient processes, and ultimately, compromised decision-making. Therefore, it is imperative to validate flat files rigorously to ensure data integrity to the uploading processes.

The importance of the data quality processes for the flat files data uploading cannot be overstated, and it encompasses several critical aspects:

1. **Accuracy:** Flat files often serve as a conduit for transferring data between systems within organizations. Ensuring data accuracy during the upload process is crucial to prevent errors that could lead to misinformation or faulty decision-making based on inaccurate data.
2. **Consistency:** Maintaining consistency in data format, structure, and content across flat files is essential for seamless integration with databases or other systems. A robust data quality process helps identify and rectify inconsistencies, ensuring that the uploaded data aligns with predefined standards.
3. **Completeness:** Flat files must contain all the necessary data fields required for their intended purpose. A data quality process helps validate the completeness of uploaded files, flagging any missing or incomplete information that could hinder downstream processes or analyses.
4. **Data Integrity:** Preserving data integrity is paramount, especially in scenarios where flat files undergo multiple transformations or manipulations before reaching their final destination. By enforcing validation checks and data integrity measures during upload, organizations can safeguard against data corruption or tampering.

4. Flat Files Uploading Process - Overview



5. Flat Files Uploading Process - Steps

1. **Automatic Process (ETL - Talend)**
 - a. The ETL process retrieves the file from a specific folder.
 - b. Performs the validations (range of values, data type, whether filling is mandatory or not, etc)
 - c. If the previous step is **OK**, the process loads the information into GCP. If something is **not OK**, the process loads the correct rows and generates an errors file containing all the incorrect ones, which need to be manually corrected by the Data Owner (sent by email).
2. **Manual Process (Data Owner - Errors File received by email)**
 - a. The Data Owner should manually correct all the incorrect rows to be uploaded in the next process run (the errors file contains detailed error descriptions to assist the Data Owner during the correction process).

6. Guidelines for Validation

- **Preparation:** Ensure that flat files conform to predefined formatting standards
- **Standardization:** Ensure consistency in naming conventions, data formats, and coding standards.
- **Data Profiling:** Analyze the structure and content of flat files to identify anomalies or inconsistencies.
- **Validation Rules:** Define clear and comprehensive validation rules tailored to specific data attributes and business requirements.
- **Automation:** Leverage automation tools and scripts to streamline the validation process and minimize manual intervention.
- **Documentation:** Maintain thorough documentation of validation procedures, including assumptions, methodologies, and outcomes.
- **Collaboration:** Foster collaboration between data analysts, domain experts, and IT professionals to address complex validation challenges effectively.

7. Resources

- **Validation Tools:** Talend ETL
- **Team:** Data Governance Squad and Data Engineering Team

8. Feedback and Suggestions

Your feedback is valuable in improving the effectiveness and efficiency of our flat files validation process. If you have any suggestions or recommendations, please feel free to share them with us.

Thank you for your commitment to maintaining data quality standards through diligent flat files validation. Let's work together to ensure the accuracy and reliability of our data assets.