

Package: `pacta.data.preparation` (via `r-universe`)

September 9, 2024

Title Prepare Data for PACTA for Investors

Version 0.1.0.9003

Description This package provides tools to prepare input datasets to be run in the `PACTA_analysis` tool.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Imports `cli` ($\geq 3.5.0$), `countrycode`, `data.table`, `devtools`, `dplyr` ($\geq 1.1.0$), `dtplyr`, `jsonlite`, `lubridate`, `rlang`, `stats`, `stringr`, `tidyr`, `tidyselect` ($\geq 1.2.0$), `withr`

Depends `R` ($\geq 3.4.0$)

LazyData true

Suggests `here`, `knitr`, `pkgdown`, `readr`, `readxl`, `rmarkdown`, `rstudioapi`, `testthat` ($\geq 3.0.0$)

Config/testthat/edition 3

Config/Needs/website `rmi-pacta/pacta.pkgdown.rmitemplate`

URL <https://github.com/RMI-PACTA/pacta.data.preparation>

BugReports <https://github.com/RMI-PACTA/pacta.data.preparation/issues>

VignetteBuilder `knitr`

Repository <https://rmi-pacta.r-universe.dev>

RemoteUrl <https://github.com/rmi-pacta/pacta.data.preparation>

RemoteRef HEAD

RemoteSha a46b90423eba350d135cdb9e044815e343bc045e

Contents

calc_shares_all_classes	2
dataprep_abcd_scen_connection	3
determine_relevant_years	4
import_ar_advanced_company_indicators	5
import_ar_masterdata	6
prepare_abcd_flags_bonds	6
prepare_abcd_flags_equity	7
prepare_ar_company_id_country_of_domicile	8
prepare_ar_company_id_credit_parent_ar_company_id	8
prepare_ar_company_id_sectors_with_assets_debt	9
prepare_ar_company_id_sectors_with_assets_ownership	9
prepare_company_id_creditor_company_id	10
prepare_entity_info	10
prepare_factset_entity_id_ar_company_id	11
prepare_factset_entity_id_credit_parent_id	12
prepare_factset_entity_id_security_mapped_sector	12
prepare_financial_data	13
prepare_fund_data	13
prepare_isin_to_fund_table	14
prepare_iss_average_sector_emission_intensities	14
prepare_iss_company_emissions	15
prepare_iss_entity_emission_intensities	16
prepare_masterdata	16
prepare_masterdata_debt	17
prepare_total_fund_list	18
quarter_to_factset_timestamp	19
quarter_to_imf_timestamp	19
standardize_asset_type_names	20
write_manifest	20
Index	21

calc_shares_all_classes

Calculate the number of shares across all equity asset classes

Description

calc_shares_all_classes() calculates the number of shares across all equity asset classes. The asset class is defined by the value in asset_type_id, and only asset classes included in include_types are summed.

calc_shares_all_classes() assumes that all input values are from the same entity/company, so a single summed value is returned. If it is desired to calculate the shares in all classes for multiple entities at the same time, a typical usage would be to apply this function "by" some entity id, e.g.

```
mutate(data, shares_all_classes = calc_shares_all_classes(shares_outstanding, asset_type_id),
       .by = "entity_id")
```

For more details see vignette("share_ownership_methodology").

Usage

```
calc_shares_all_classes(
  shares_outstanding,
  asset_type_id,
  include_types = c("CP", "EQ", "PF")
)
```

Arguments

`shares_outstanding` (numeric vector) contains the number of shares outstanding

`asset_type_id` (character vector) contains an the asset type identifier

`include_types` (character vector) defines which of the `asset_type_id` values should be included in the sum

Value

a numeric value containing the number of shares across all equity asset classes

dataprep_abcd_scen_connection

Combine ABCD and scenario data into the [equity/bonds]_abcd_scenario.rds format that is used by portfolio.analysis

Description

Combine ABCD and scenario data into the [equity/bonds]_abcd_scenario.rds format that is used by portfolio.analysis

Usage

```
dataprep_abcd_scen_connection(
  abcd_data,
  scenario_data,
  reference_year,
  relevant_years,
  tech_exclude,
  scenario_geographies_list,
  sector_list,
  other_sector_list,
  global_aggregate_scenario_sources_list,
```

```

    global_aggregate_sector_list,
    scenario_regions,
    index_regions
  )

```

Arguments

abcd_data A tibble containing the ABCD data
scenario_data A tibble containing the scenario data
reference_year A single numeric specifying the market share target reference year
relevant_years A numeric vector containing all relevant years to be calculated
tech_exclude A character vector containing the technologies to be excluded
scenario_geographies_list
 A character vector containing the scenario geographies to be used
sector_list A character vector containing the sectors to be included
other_sector_list
 A character vector containing the sectors considered "other"
global_aggregate_scenario_sources_list
 A character vector containing the scenario sources to be included in the global
 aggregate
global_aggregate_sector_list
 A character vector containing the sectors to be included in the global aggregate
scenario_regions
 A character vector containing the scenario regions
index_regions A character vector containing the index regions

Value

A tibble with the combined ABCD and scenario data

```

determine_relevant_years
  Determine relevant years

```

Description

Determine relevant years

Usage

```
determine_relevant_years(market_share_target_reference_year, time_horizon)
```

Arguments

market_share_target_reference_year A single numeric value determining the Market Share target reference year

time_horizon A single numeric value determining the number of forward looking years

Value

A numeric vector containing all of the relevant years

import_ar_advanced_company_indicators

Import the data from a version of Asset Resolution's proprietary Advanced Company Indicators XLSX into a tidy data frame.

Description

Import the data from a version of Asset Resolution's proprietary Advanced Company Indicators XLSX into a tidy data frame.

Usage

```
import_ar_advanced_company_indicators(
  filepath,
  drop_nas = TRUE,
  fix_names = FALSE,
  as_factor = TRUE
)
```

Arguments

filepath Path to the XLSX file.

drop_nas A logical indicating whether rows with an NA value after pivoting to long-format should be dropped (default is TRUE).

fix_names A logical indicating whether the column names should be fixed to snakecase format. (e.g. Company Name becomes company_name). By default, column names are not changed (i.e. FALSE).

as_factor A logical indicating whether the character columns should be converted to factors (default is TRUE).

Value

A tibble including all the data from the "Company Information", "Company ISINs", "Company Emissions", and "Company Activities" tabs combined into one tidy tibble.

```
import_ar_masterdata Import the data from a version of Asset Resolution's bespoke
masterdata_*.csv files into a tidy data frame.
```

Description

Import the data from a version of Asset Resolution's bespoke masterdata_*.csv files into a tidy data frame.

Usage

```
import_ar_masterdata(filepath, drop_nas = TRUE, id_as_string = FALSE)
```

Arguments

filepath	Path to the CSV file.
drop_nas	A logical indicating whether rows with an NA value after pivoting to long-format should be dropped (default is TRUE).
id_as_string	A logical indicating whether the company_id column should be imported as a character vector. By default, the company_id column is imported as a numeric (i.e. FALSE).

Value

A tidy, long-format tibble of all the data in the masterdata_*.csv file with an added consolidation_method column to record which of "ownership" or "debt" file was imported.

```
prepare_abcd_flags_bonds
Title
```

Description

Title

Usage

```
prepare_abcd_flags_bonds(
  financial_data,
  factset_entity_id__ar_company_id,
  factset_entity_id__security_mapped_sector,
  ar_company_id__sectors_with_assets__debt,
  factset_entity_id__credit_parent_id
)
```

Arguments

financial_data A data frame containing financial data

factset_entity_id__ar_company_id
A data frame containing a factset_entity_id to ar_company_id look up table

factset_entity_id__security_mapped_sector
A data frame containing a factset_entity_id to security_mapped_sector look up table

ar_company_id__sectors_with_assets__debt
A data frame containing a ar_company_id to sectors_with_assets look up table for debt

factset_entity_id__credit_parent_id
A data frame containing a factset_entity_id to credit_parent_id look up table

Value

A data frame

prepare_abcd_flags_equity
Title

Description

Title

Usage

```
prepare_abcd_flags_equity(
  financial_data,
  factset_entity_id__ar_company_id,
  factset_entity_id__security_mapped_sector,
  ar_company_id__sectors_with_assets__ownership
)
```

Arguments

financial_data A data frame containing financial data

factset_entity_id__ar_company_id
A data frame containing a factset_entity_id to ar_company_id look up table

factset_entity_id__security_mapped_sector
A data frame containing a factset_entity_id to security_mapped_sector look up table

ar_company_id__sectors_with_assets__ownership
A data frame containing a ar_company_id to sectors_with_assets look up table for ownership

Value

A data frame

`prepare_ar_company_id__country_of_domicile`

Prepare an ar_company_id__country_of_domicile lookup table from the entity_info data

Description

Prepare an ar_company_id__country_of_domicile lookup table from the entity_info data

Usage

```
prepare_ar_company_id__country_of_domicile(entity_info)
```

Arguments

`entity_info` A data frame containing the entity info

Value

A tibble

`prepare_ar_company_id__credit_parent_ar_company_id`

Prepare an ar_company_id__credit_parent_ar_company_id lookup table from the entity_info data

Description

Prepare an ar_company_id__credit_parent_ar_company_id lookup table from the entity_info data

Usage

```
prepare_ar_company_id__credit_parent_ar_company_id(entity_info)
```

Arguments

`entity_info` A data frame containing the entity info

Value

A tibble

```
prepare_ar_company_id__sectors_with_assets__debt
  Prepare an ar_company_id__sectors_with_assets__debt lookup
  table from the masterdata_debt_datastore data
```

Description

Prepare an ar_company_id__sectors_with_assets__debt lookup table from the masterdata_debt_datastore data

Usage

```
prepare_ar_company_id__sectors_with_assets__debt(  
  masterdata_debt_datastore,  
  relevant_years  
)
```

Arguments

masterdata_debt_datastore A data frame containing processed production data from Asset Impact's masterdata_ownership CSV

relevant_years A numeric vector containing the relevant years of data to include

Value

A tibble

```
prepare_ar_company_id__sectors_with_assets__ownership
  Prepare an ar_company_id__sectors_with_assets__ownership
  lookup table from the masterdata_ownership_datastore data
```

Description

Prepare an ar_company_id__sectors_with_assets__ownership lookup table from the masterdata_ownership_datastore data

Usage

```
prepare_ar_company_id__sectors_with_assets__ownership(  
  masterdata_ownership_datastore,  
  relevant_years  
)
```

Arguments

masterdata_ownership_datastore
 A data frame containing processed production data from Asset Impact's masterdata_ownership CSV

relevant_years A numeric vector containing the relevant years of data to include

Value

A tibble

prepare_company_id__creditor_company_id
Prepare a company_id__creditor_company_id lookup table from Asset Impact's masterdata_debt data

Description

Prepare a company_id__creditor_company_id lookup table from Asset Impact's masterdata_debt data

Usage

```
prepare_company_id__creditor_company_id(masterdata_debt)
```

Arguments

masterdata_debt
 A data frame containing raw production data from Asset Impact's masterdata_debt CSV

Value

A tibble

prepare_entity_info *Prepare the entity_info output data frame from data frames imported from the factset_entity_info.rds and ar_company_id__factset_entity_id.rds files*

Description

Prepare the entity_info output data frame from data frames imported from the factset_entity_info.rds and ar_company_id__factset_entity_id.rds files

Usage

```
prepare_entity_info(
  data,
  factset_entity_id__ar_company_id,
  factset_industry_map_bridge,
  factset_manual_pacta_sector_override
)
```

Arguments

`data` A data frame containing the imported factset_entity_info.rds file

`factset_entity_id__ar_company_id` A data frame containing the imported ar_company_id__factset_entity_id.rds file with columns factset_entity_id and ar_company_id

`factset_industry_map_bridge` A data frame containing the imported factset_industry_map_bridge.rds

`factset_manual_pacta_sector_override` A data frame containing the imported factset_manual_pacta_sector_override

Value

A tibble properly prepared to be saved as the entity_info.rds output file

```
prepare_factset_entity_id__ar_company_id
```

Prepare a factset_entity_id__ar_company_id lookup table from Asset Impact's ar_company_id__factset_entity_id crosswalk

Description

Prepare a factset_entity_id__ar_company_id lookup table from Asset Impact's ar_company_id__factset_entity_id crosswalk

Usage

```
prepare_factset_entity_id__ar_company_id(ar_company_id__factset_entity_id)
```

Arguments

`ar_company_id__factset_entity_id` A data frame containing a production data company ID to financial data entity lookup table

Value

A tibble

```
prepare_factset_entity_id__credit_parent_id  
  Prepare a factset_entity_id__credit_parent_id lookup table  
  from entity_info
```

Description

Prepare a factset_entity_id__credit_parent_id lookup table from entity_info

Usage

```
prepare_factset_entity_id__credit_parent_id(entity_info)
```

Arguments

entity_info A data frame containing the entity info

Value

A tibble

```
prepare_factset_entity_id__security_mapped_sector  
  Prepare a factset_entity_id__security_mapped_sector lookup  
  table from entity_info
```

Description

Prepare a factset_entity_id__security_mapped_sector lookup table from entity_info

Usage

```
prepare_factset_entity_id__security_mapped_sector(entity_info)
```

Arguments

entity_info A data frame containing the entity info

Value

A tibble

```
prepare_financial_data
```

Prepare the financial_data output data frame from the imported factset_financial_data.rds file

Description

Prepare the financial_data output data frame from the imported factset_financial_data.rds file

Usage

```
prepare_financial_data(data, issue_code_bridge)
```

Arguments

data A data frame containing the imported factset_financial_data.rds file

issue_code_bridge

A data frame containing data that bridges from factset issue codes to one of c("Listed Equity", "Corporate Bond", "Fund", "Other")

Value

A tibble properly prepared to be saved as the financial_data.rds output file

```
prepare_fund_data
```

Prepare fund data, filtering to funds with data according to a given threshold and adding a MISSINGWEIGHT holding for the difference

Description

Prepare fund data, filtering to funds with data according to a given threshold and adding a MISSINGWEIGHT holding for the difference

Usage

```
prepare_fund_data(fund_data, threshold = 0)
```

Arguments

fund_data A data frame containing fund data

threshold A numeric value between 0 and 1 (inclusive) indicating the allowable percentage of the total fund value that the summed values of its component holdings should be equal to or greater than

Value

A tibble

prepare_isin_to_fund_table

Prepare isin_to_fund_table, filtering out fsyms that have more than 1 row and either no fund data or fund data for both rows

Description

Prepare isin_to_fund_table, filtering out fsyms that have more than 1 row and either no fund data or fund data for both rows

Usage

```
prepare_isin_to_fund_table(isin_to_fund_table, fund_data)
```

Arguments

isin_to_fund_table

A data frame containing isin_to_fund_table data

fund_data

A data frame containing fund data

Value

A tibble

prepare_iss_average_sector_emission_intensities

Prepare a iss_average_sector_emission_intensities object

Description

Prepare a iss_average_sector_emission_intensities object

Usage

```
prepare_iss_average_sector_emission_intensities(  
  iss_company_emissions,  
  factset_financial_data,  
  factset_entity_info,  
  factset_entity_financing_data,  
  currencies  
)
```

Arguments

iss_company_emissions
A data frame containing iss_company_emissions data

factset_financial_data
A data frame containing factset_financial_data data

factset_entity_info
A data frame containing factset_entity_info data

factset_entity_financing_data
A data frame containing factset_entity_financing_data data

currencies
A data frame containing currency exchange rate data

Value

A data frame containing the prepared iss_average_sector_emission_intensities object

```
prepare_iss_company_emissions
  Prepare an iss_company_emissions object from
  factset_iss_emissions_data
```

Description

Prepare an iss_company_emissions object from factset_iss_emissions_data

Usage

```
prepare_iss_company_emissions(factset_iss_emissions_data)
```

Arguments

factset_iss_emissions_data
A data frame containing ISS emissions data

Value

A tibble

```
prepare_iss_entity_emission_intensities
    Prepare a iss_entity_emission_intensities object
```

Description

Prepare a iss_entity_emission_intensities object

Usage

```
prepare_iss_entity_emission_intensities(
  iss_company_emissions,
  factset_financial_data,
  factset_entity_info,
  factset_entity_financing_data,
  currencies
)
```

Arguments

```
iss_company_emissions    A data frame containing iss_company_emissions data
factset_financial_data   A data frame containing factset_financial_data data
factset_entity_info      A data frame containing factset_entity_info data
factset_entity_financing_data A data frame containing factset_entity_financing_data data
currencies               A data frame containing currency exchange rate data
```

Value

A data frame containing the prepared iss_entity_emission_intensities object

```
prepare_masterdata    Prepare the masterdata_ownership_datastore or
masterdata_debt_datastore output data frame from an im-
port of a raw AR masterdata_* CSV
```

Description

Prepare the masterdata_ownership_datastore or masterdata_debt_datastore output data frame from an import of a raw AR masterdata_* CSV

Usage

```
prepare_masterdata(
  data,
  ar_company_id__country_of_domicile,
  pacta_financial_timestamp,
  zero_emission_factor_techs
)
```

Arguments

`data` A dataframe containing the raw input of an AR masterdata_* CSV files

`ar_company_id__country_of_domicile` A data frame with two columns mapping `ar_company_id` to `country_of_domicile`

`pacta_financial_timestamp` A single element character vector specifying the timestamp in the PACTA format, e.g. "2021Q4"

`zero_emission_factor_techs` A character vector listing technologies that will have emission factors manually forced to 0

Value

A tibble properly prepared to be saved as the `masterdata_ownership_datastore.rds` or `masterdata_debt_datastore.rds` output file

```
prepare_masterdata_debt
```

Prepare the masterdata_debt_datastore object from a raw masterdata_debt CSV

Description

Prepare the `masterdata_debt_datastore` object from a raw `masterdata_debt` CSV

Usage

```
prepare_masterdata_debt(
  masterdata_debt_raw,
  ar_company_id__country_of_domicile,
  ar_company_id__credit_parent_ar_company_id,
  pacta_financial_timestamp,
  zero_emission_factor_techs
)
```

Arguments

`masterdata_debt_raw`
A data frame containing the raw data from a `masterdata_debt` CSV

`ar_company_id__country_of_domicile`
A data frame containing an `ar_company_id` to `country_of_domicile` lookup

`ar_company_id__credit_parent_ar_company_id`
A data frame containing an `ar_company_id` to `credit_parent_ar_company_id` lookup

`pacta_financial_timestamp`
A single character vector containing the PACTA financial timestamp, e.g. 2023Q4

`zero_emission_factor_techs`
A character vector containing the zero emission factor technologies

Value

A data frame containing the prepared `masterdata_debt_datastore`

`prepare_total_fund_list`

Prepare a `total_fund_list` object from `fund_data`

Description

Prepare a `total_fund_list` object from `fund_data`

Usage

```
prepare_total_fund_list(fund_data)
```

Arguments

`fund_data` A data frame containing fund data

Value

A tibble

quarter_to_factset_timestamp

Convert a PACTA style quarter string to a FactSet style date string for the last day of that quarter

Description

Convert a PACTA style quarter string to a FactSet style date string for the last day of that quarter

Usage

quarter_to_factset_timestamp(quarter)

Arguments

quarter A character vector containing PACTA style quarter strings in the form e.g. "2021Q4"

Value

A character vector containing the equivalent FactSet style date strings for the last day of the quarter e.g. "2021-12-31"

quarter_to_imf_timestamp

Convert a PACTA style quarter string to an IMF style quarter string

Description

Convert a PACTA style quarter string to an IMF style quarter string

Usage

quarter_to_imf_timestamp(quarter)

Arguments

quarter A character vector containing PACTA style quarter strings in the form e.g. "2021Q4"

Value

A character vector containing the equivalent IMF style quarter strings e.g. "2021-Q4"

standardize_asset_type_names	<i>Standardize asset type names</i>
------------------------------	-------------------------------------

Description

Standardize asset type names

Usage

```
standardize_asset_type_names(factset_issue_code_bridge)
```

Arguments

factset_issue_code_bridge	A data frame containing the FactSet issue code bridge
---------------------------	---

Value

A data frame containing a FactSet issue_type_code to asset_type lookup

write_manifest	<i>Write a manifest.json file to the specified path including critical information about the files and parameters used to prepare the data</i>
----------------	--

Description

Write a manifest.json file to the specified path including critical information about the files and parameters used to prepare the data

Usage

```
write_manifest(path, parameters, input_files, output_files)
```

Arguments

path	A single string specifying a filepath to save the JSON file to
parameters	A list containing all parameters used to create the data
input_files	A vector with filepaths of input files used to create the output data.
output_files	A vector with filepaths of output files created.

Value

Called for the side-effect of writing a JSON file to disk

Index

calc_shares_all_classes, [2](#)

dataprep_abcd_scen_connection, [3](#)

determine_relevant_years, [4](#)

import_ar_advanced_company_indicators, [5](#)

import_ar_masterdata, [6](#)

prepare_abcd_flags_bonds, [6](#)

prepare_abcd_flags_equity, [7](#)

prepare_ar_company_id__country_of_domicile, [8](#)

prepare_ar_company_id__credit_parent_ar_company_id, [8](#)

prepare_ar_company_id__sectors_with_assets__debt, [9](#)

prepare_ar_company_id__sectors_with_assets__ownership, [9](#)

prepare_company_id__creditor_company_id, [10](#)

prepare_entity_info, [10](#)

prepare_factset_entity_id__ar_company_id, [11](#)

prepare_factset_entity_id__credit_parent_id, [12](#)

prepare_factset_entity_id__security_mapped_sector, [12](#)

prepare_financial_data, [13](#)

prepare_fund_data, [13](#)

prepare_isin_to_fund_table, [14](#)

prepare_iss_average_sector_emission_intensities, [14](#)

prepare_iss_company_emissions, [15](#)

prepare_iss_entity_emission_intensities, [16](#)

prepare_masterdata, [16](#)

prepare_masterdata_debt, [17](#)

prepare_total_fund_list, [18](#)

quarter_to_factset_timestamp, [19](#)

quarter_to_imf_timestamp, [19](#)

standardize_asset_type_names, [20](#)

write_manifest, [20](#)