Digiwhist sustainability - update 4 release notes

Structured data

Majority of time was spent on fixing and improving of process of retrieving structured data from unstructured data and it's correct cleaning for following sources:

- Romania http://data.gov.ro/
 - Performance issues that avoided all data to be processed during the update 3 were fixed and currently all data from the has been processed which is 19.363.574 individual publications resulting into 18.571.019 contracting processes. Most of them are direct purchases.
 - Also some minor issues were fixed
- Portugal http://www.base.gov.pt
 - Major improvements in handling of multi contract procurements. Each contract is now stored as a separate lot
- Austria https://www.data.gv.at
 - o Until now only TED data were available for Austria.
 - Complete set of workers for processing Austrian national source were implemented
 - Total number of contracting processes obtained from this source is 25.803
- Spain https://www.hacienda.gob.es
 - Complete set of workers for new Spanish open data source has been implemented
 - Total number of contracting processes obtained from this source is 954.136
 - Because its data overlaps with http://contrataciondelestado.es a deduplication mechanism needed to be implemented
- Czechia https://www.vestnikverejnychzakazek.cz
 - Major methodological improvements focused on framework agreement handling and DPS tenders processing
- Slovakia http://www.uvo.gov.sk
 - Major methodological improvements focused on framework agreement handling and DPS tenders processing
- Ireland
 - minor fixes
- Italy https://dati.anticorruzione.it
 - Complete set of workers for new Italian open data source has been implemented
 - Total number of contracting processes obtained from this source is 14.079.312

Indicators

Several new indicators were implemented

Lot level indicators

Indicators relevant to each lot of a tender

- TRANSPARENCY_BIDDER_NAME_MISSING Share of non-missing bidder name
- TRANSPARENCY_BIDDER_ID_MISSING Share of non-missing bidder id
- TRANSPARENCY_VALUE_MISSING Share of non-missing contract value
- INTEGRITY_WINNER_CA_SHARE Share of expenses paid by buyer to a specific supplier
- INTEGRITY COST OVERRUN Cost overrun

Tender level indicators

Indicators relevant to a whole contracting process

- TRANSPARENCY BUYER NAME MISSING Share of non-missing buyer name
- TRANSPARENCY_YEAR_MISSING Share of non-missing year info
- TRANSPARENCY_TITLE_MISSING Share of non-missing title
- TRANSPARENCY_MARKET_MISSING Share of non-missing product type
- TRANSPARENCY_BUYER_LOC_MISSING Share of non-missing buyer location
- TRANSPARENCY_PROC_METHOD_MISSING Share of non-missing procedure type
- TRANSPARENCY_AWARD_DATE_MISSING Share of non-missing award date

Opentender

- Updated with the new version of data
- Download page was completely restructured https://opentender.eu/hu/download
- TED only dataset was added to a download page

CSV export

• all new indicators were added to the CSV export

Data update

Links to updated CSV data

https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/AM_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/AT_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/BE_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/BG_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/CH_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/CY_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/CZ_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/DE_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/DE_data.csv

https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/DK data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/EE_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/ES data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/FI data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/FR_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/GE data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/GR_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/HR data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/HU data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/IE_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/IS data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/IT_data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/LT data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/LU data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/LV data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/MT data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/NL data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/NO data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/PL data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/PT data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/RO data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/RS data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/SE data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/SI data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/SK data.csv https://s3.eu-central-1.amazonaws.com/digiwhist-data/CSV/UK data.csv

Links to updated Digiwhist JSON data

https://digiwhist-data.s3.eu-central-1.amazonaws.com/D2 6/tender/AM data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/AT data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/BE data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/BG data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/CH data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/CY data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/CZ data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/DE_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/DK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/EE data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/ES data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/FI data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/FR_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/GE data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/GR data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/HR data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/HU_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/IE_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/IS_data.json.tar.gz

https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/IT_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/LU_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/LV_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/MT_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/ML_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/NL_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/PL_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/PT_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/RO_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/RS_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SE_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SI_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_data.json.tar.gz

Links to updated OCDS data

https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/AM ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/AT ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/BE ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/BG_ocds_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/CH ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/CY ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/CZ ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/DE ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/DK ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/EE ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/ES ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/FI ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/FR ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/GE ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/GR ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/HR ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/HU ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/IE ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/IS ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/IT ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/LT ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/LU ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/LV ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/MT ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/NL ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/NO_ocds_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/PL ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/PT ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/RO ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/RS_ocds_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SE_ocds_data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2 6/tender/SI ocds data.json.tar.gz https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/SK_ocds_data.json.tar.gz

https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_6/tender/UK_ocds_data.json.tar.gz

Validation reports

Current version of validation reports can be found at:

https://github.com/digiwhist/wp2_documents/tree/master/validation_reports