

**EU Grant Agreement number: 645852**

**Project acronym: DIGIWHIST**

**Project title: The Digital Whistleblower: Fiscal Transparency, Risk  
Assessment and the Impact of Good Governance Policies  
Assessed**

**Work Package: 2 – Quantitative Data Collection and Cleaning**

**Title of deliverable: D2.5 Cleaned and Structured Databases**

Due date of deliverable: 31/05/2017

Actual submission date: 31/05/2017

Authors: Jan Hrubý, Tomáš Pošepný, Jakub Kraška, Tomáš Mrázek,  
Marek Mikeš (UCAM) and Jiří Skuhrovec (Datlab)

Organization name of lead beneficiary for this deliverable:  
University of Cambridge

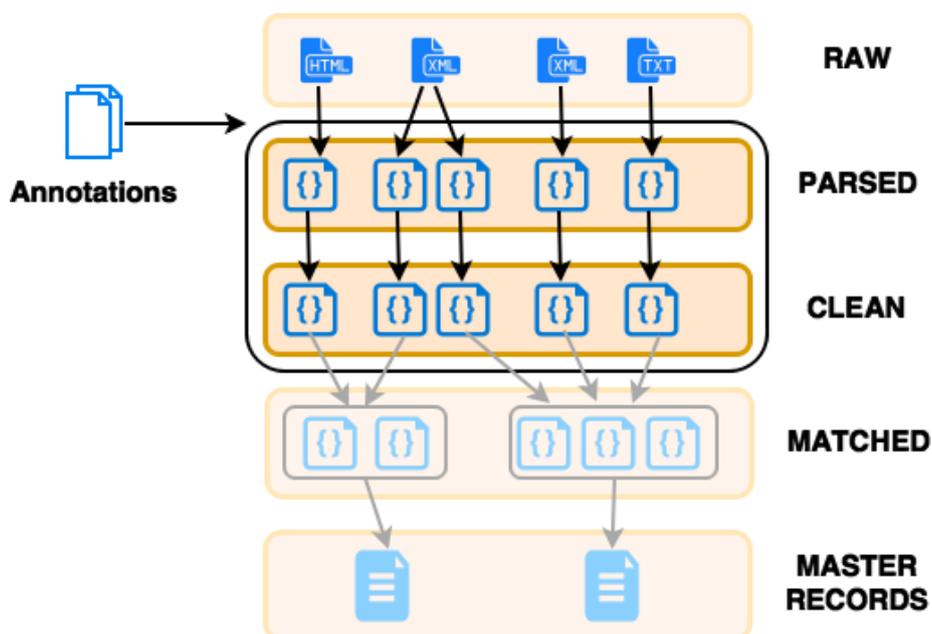
<b>Dissemination Level</b>		
<b>PU</b>	Public	X
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>Co</b>	Confidential, only for members of the consortium (including the Commission Services)	

# Introduction

The collection of raw data (D2.4) is about understanding source systems, what data they offer and how the data can be obtained from a source; to create a structured database we need to understand the data itself and store it according to a data template that has been designed to best support analytical work in other deliverables of the DIGIWHIST project.

The system for data collection and transformation has been designed to process data in several stages so that we can go back a step at any time without losing any information. This enables us to identify the exact point in the process where errors appear and fix them without having to repeat full (and costly) data extraction processes. In short, it makes the whole development-validation cycle more efficient.

This deliverable describes how we treat data in the second (parsing) and third (cleaning) stages.



## Approach

### Structuring data

Converting unstructured data into structured data requires a combination of knowledge from public procurement experts and work by software developers who convert such knowledge into executable programs. In this phase of data processing the only action is to extract information directly, in the same format as in the source. During this data parsing phase, there is no computation creating a new piece of information. This topic is covered by a wide range of academic work, starting from pioneering works like Frakes and Yates (1992) or Hammer et. al. (1997). Initially, our work followed modern machine learning approaches such as those described in Lindes et. al. (2015) or perhaps more notably Svátek et. al.

(2015), who develop and pioneer ontology-based data handling methods specifically in public procurement. Yet in the end, more explicit and tailored approaches, such as those described in Embley et. al. (2016) appeared to be more reliable.

## Expert knowledge

As a first step in raw data processing we need to identify a piece of information which is reasonable to store separately as a single record. After that, experts can classify such publications and describe a sample representative for each class. Such descriptions can have various formats. We always try to choose a format which is most appropriate for a source's specificities and understandable for developers. Even a description of structured data (XML files) can differ depending on the sources so there is no single method providing guidance on each source description. We mostly use three different techniques:

- XSD

If the source data are defined by an XML schema definition, annotations can be done by adding comments to XSD file

```
▼<xs:element name="LOTS" minOccurs="0">
  ▼<xs:complexType>
    ▼<xs:sequence maxOccurs="unbounded">
      ▼<xs:element name="LOT">
        ▼<xs:complexType>
          ▼<xs:sequence>
            <!-- lot.lotNumber -->
            <xs:element name="NUM" type="xs:string" minOccurs="0"/>
            <!-- lot.title -->
            <xs:element name="INTITULE" type="xs:string" minOccurs="0"/>
            <!-- lot.description -->
            <xs:element name="DESCRIPTION" type="xs:string" minOccurs="0"/>
            <!-- lot.CPVs -->
            ▼<xs:element name="CPV" minOccurs="0" maxOccurs="unbounded">
              ▼<xs:complexType>
                ▼<xs:sequence>
                  <!-- lot.CPVs main cpv -->
                  <xs:element name="PRINCIPAL" type="typeCPVBasePrincipale"/>
                  <!-- lot.CPVs additional cpv -->
                  <xs:element name="SUPPLEMENTAIRE" type="typeCPVBaseSupplementaire" minOccurs="0" maxOccurs="unbounded"/>
                </xs:sequence>
              </xs:complexType>
            </xs:element>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

- Element listing

For example TED provides element listing in the form of an Excel spreadsheet. In such a case we complete such a listing by mapping each element to our data template

1	XMI element	Label_key	DIGIWHIST DB correspondence	Comment
50	TYPE_SUPPLIES_CONTRACT VALUE="RENTAL"	rental	-	
51	TYPE_SUPPLIES_CONTRACT VALUE="HIREPURCHASE"	hirepurchase	-	
52	TYPE_SUPPLIES_CONTRACT VALUE="COMBINATION"	combination	-	
53	TYPE_CONTRACT VALUE="SERVICES"	services	tender.supplyType	
54	SERVICE_CATEGORY_NO	service_cat_no	-	
55		agree_pub_cat_17_27_c2_no	-	
56		agree_pub_cat_17_27_c2_yes	-	
57	LOCATION_NUTS/LOCATION	mainsiteplace_works_delivery	tender.addressOfImplementation.rawAddress	
58	NUTS CODE=	nutscode	tender.addressOfImplementation.nuts	
59	F06_NOTICE_INVOLVES	framework_dps_info	-	
60	CONCLUSION_FRAMEWORK_AGREEMENT	notice_involves_framework	tender.isFrameworkAgreement	
61	CONTRACTS_DPS	notice_involves_contracts_dps	tender.isDPS	
62	SHORT_CONTRACT_DESCRIPTION	descr_short_contract_purch	tender.description	
63	CPV	cpv	-	
64	CPV_MAIN	cpv_main	tender.CPVs.code	tender.CPVs.isMain="true"
65	CPV_CODE CODE=	cpv_main_object	tender.CPVs.code	tender.CPVs.isMain="true"
66	CPV_SUPPLEMENTARY_CODE CODE=	cpv_supplem	tender.CPVs.code	tender.CPVs.isMain="false"
67	CPV_ADDITIONAL	cpv_supplem	tender.CPVs.code	tender.CPVs.isMain="false"
68	CPV_CODE CODE=	cpv_supplem	tender.CPVs.code	tender.CPVs.isMain="false"
69	CPV_SUPPLEMENTARY_CODE CODE=	cpv_supplem	tender.CPVs.code	tender.CPVs.isMain="false"
70	CONTRACT_COVERED_GPA VALUE="NO"	gpa_info:gpa_covered_no	tender.isCoveredByGPA="false"	
71	CONTRACT_COVERED_GPA VALUE="YES"	gpa_info:gpa_covered_yes	tender.isCoveredByGPA="true"	
72	TOTAL_FINAL_VALUE	value_total_final	tender.finalPrice.netAmount	
73	COSTS_RANGE_AND_CURRENCY_WITH_VAT_RATE CURRENCY=		-	
74	VALUE_COST	value; currency	tender.finalPrice.currency	
75	RANGE_VALUE_COST		-	
76	LOW_VALUE	lowest_offer	tender.finalPrice.minNetAmount	
77	HIGH_VALUE	highest_offer	tender.finalPrice.maxNetAmount	
78	EXCLUDING_VAT	excl_vat	tender.finalPrice.vat	C72,C76, C77 is net if this it true
79	INCLUDING_VAT	incl_vat	tender.finalPrice.vat	C72,C76, C77 is gross if this it true

- Visual highlighting

Sometimes, even if the data are structured, we are not able to identify any definition of such data. Therefore, we annotate such source like it is an unstructured source by using visual highlighting

```
<FORM_SECTION> tender.publication[1].formType
<F02_2014 FORM="F02" CATEGORY="ORIGINAL" LG="NB">
```

```
<CONTRACTING_BODY>
  <ADDRESS_CONTRACTING_BODY> tender.buyer
  <OFFICIALNAME>Sykehusinnkj p HF</OFFICIALNAME> name
  <NATIONALID>916 879 067</NATIONALID> bodylds.id bodylds.type=national
  <ADDRESS>Hermetikken Tollbugata 7</ADDRESS> address.street
  <TOWN>VADS </TOWN> address.town
  <POSTAL_CODE>9800</POSTAL_CODE> address.postcode
  <COUNTRY VALUE="NO"/> address.country
  <CONTACT_POINT>Sigrid Elisabeth Kordt</CONTACT_POINT> contractPoint
  <PHONE>+47 92083484</PHONE> phone
  <E_MAIL>elisabeth.kordt@sykehusinnkjop.no</E_MAIL> email
  <LOCATION code="000000"/> address.state -> this is a county code - possibly useful for calculating NUTS codes
  <URL_GENERAL>http://sykehusinnkjop.no</URL_GENERAL> address.url
  <URL_BUYER>
    https://kgv.doffin.no/ctm/Company/CompanyInformation/Index/72439
  </URL_BUYER>
  </ADDRESS_CONTRACTING_BODY>
```

We might want to parse both url-s here - though the URL\_BUYER seems to be a central procuring body website

### II.1.1 Naslov

Naslov: Poslovni najem vozil z nizkimi emisijami tender.title  
Referenčna številka dokumenta: 000057/2017 publication.buyerAssignedId

### II.1.2 Glavna koda CPV

Glavni besednjak      Dopolnilni besednjak

34110000  
CPV.code isMain=yes

### II.1.3 Vrsta naročila

Storitve tender.supplyType

### II.1.4 Kratek opis

Poslovni najem vozil z nizkimi emisijami za potrebe naročnika. tender.description

### II.1.5 Ocenjena skupna vrednost

Vrednost brez DDV: 1.006.200,00 EUR tender.estimatedPrice      "brez DDV": excluding VAT

### II.1.6 Informacije o sklopih

## Parsers

In the next step, this conceptual knowledge has to be converted into executable programs.

1. Time estimates for each template are created
2. Priorities are set for each template based on
  - a. Publication type
    - i. Contract notice, Contract award, Contract cancellation
    - ii. Contract update, Prior contract notice
    - iii. Other
  - b. Exact or estimated frequency of occurrence
3. According to the resources allocated to a particular source, parsers for most priority templates are implemented so that we obtain as much information about public procurement as possible within the limited time available.

We use standard technologies for data extraction. For structured data an approach based on XPath or CSS selectors is absolutely sufficient. For these purposes we use a Jsoup library that provides an API for extracting data using the best of DOM, CSS, and jquery-like methods.

The processing of unstructured HTML sources differs with respect to the implementation details of the source system. Some systems use easily identifiable elements and can be

handled almost as well as structured sources. On the other hand, some sources require text preprocessing (e.g. division into sections) or searching based on regular expressions which is always the decision of a developer who knows the structure of the source best and considers the advantages and disadvantages of different approaches.

## Cleaning data

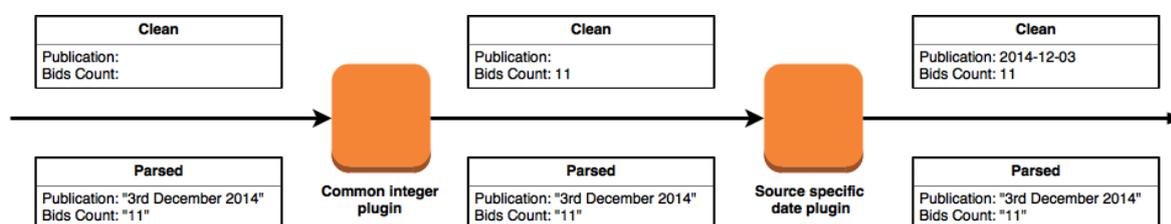
Since structured (parsed) data are stored as text values, further processing is needed to convert them into proper data types or, in the case of text values, unify them as enumeration values where possible. Additional computation based on values published in particular publications can take place at this data processing stage because some information is published indirectly. For example, combining a price inclusive of VAT with the relevant VAT rate defines the net amount even it's not directly stated.

## Cleaning plugins

Each program that processes structured data from a specific source registers a set of cleaning plugins. Such plugins are capable of cleaning one or more variables, with the program running these plugins sequentially in the order they were registered. This means, for example, that the program can use a DatePlugin (which implements the rules for processing dates) on all sources to clean all dates (e.g. award deadline, award decision date, cancellation date etc.). If some special treatment of a date for a particular source is then needed, a new plugin which overrides the original behaviour with a new one can be registered at the end of the plugin chain.

Since each plugin reads the information from a parsed record, processes it and stores it into a clean record which is empty at the beginning of the process, the program must specify how to handle each variable otherwise it won't be set in a final clean tender record.

The whole software architecture of data cleaners is described in D2.1.



## Enumeration values

For analytical purposes, the data must be as comparable as possible across countries and over time within the same country so we try to map national text values to a set of standard enumeration values. Because of the variation in sources and national legislation, this is not always feasible. Therefore, for the most important variables we will publish both the original and standardized values (the latter might be empty if we are not able to provide a mapping). There are several variables which we are trying to standardize across all sources. All mapping will be a part of the final methods paper D2.8. These variables are:

- tender.procedureType
- tender.buyer[i].buyerType
- tender.publication[i].formType
- tender.selectionMethod

## Post processing

As mentioned in the first paragraph of this chapter, the cleaning process is not only about converting existing values from their text representations into the final data type, but also about computing those pieces of information that can be constructed from the existing data contained in a single publication. These are for example:

- Computation of lot status based on publication type
- Computation of net amount by combining amount including VAT with VAT rate and vice versa
- Price conversion to EUR
- Setting award criteria to 100% price when the selection method is lowest price
- Setting NUTS value based on postcode

These and other rules are a matter of discussions between consortium members and the complete list will be described in detail in the final deliverable D2.8 methods paper.

## Results & validation

From the outset, the goal of DIGIWHIST has been to deliver as precise and complete data as possible. The UCAM team has worked in close cooperation with DIGIWHIST partner Datlab to achieve this. Datlab carries out manual checking to compare structured data with the original values published on a source. This is a basic process that gives fundamental feedback on data quality and is focused on specific errors (e.g. a missing title for a particular tender). To see how a set of errors that were discovered manually influenced the whole dataset we developed a range of statistics that enables us to compare the completeness of two versions of the same dataset. These statistics for all sources can be found in Annex 1 of this deliverable. In addition to monitoring data quality between different versions of one dataset, statistics also help us to reveal a critical part of our dataset and help us to prioritize on which sources and variables we should focus our efforts.

As an illustration, in practice, the consortium can decide that having a final price for more than 50% of tenders in Spain is more important for consecutive tasks (like indicator computation) than improving the Latvian procedure type variable which might be at a satisfactory level. This is very important because the final dataset definitely won't be 100% complete for all variables and all countries for a variety of reasons, such as a lack of information on a source, the poor structure of a source or errors in data extraction programs. Given the limited time available, we'll have to make further such decisions on how to manage this in the upcoming months.

During RP1 we devoted significant time to mapping potential sources suitable for data processing. We detected 22 sources with sufficient data scope and quality for DIGIWHIST's analytical goals, each of them geographically covering at least one whole jurisdiction. We have detailed validity statistics for these sources only, which continue to guide our decisions on the final database content which will be prepared for deliverables D2.6-2.8 and D3.6. The

table below shows the total number of publications for each jurisdiction that have been processed so far, including those which turned out to be of insufficient scope or quality for further analytical purposes.

<b>Poland</b>	3164481	<b>Belgium</b>	58860
<b>France</b>	3090667	<b>Romania</b>	50461
<b>Portugal</b>	818946	<b>Bulgaria</b>	47690
<b>Spain</b>	473675	<b>Finland</b>	37876
<b>Czech Republic</b>	396061	<b>Austria</b>	34638
<b>Germany</b>	307413	<b>Switzerland</b>	32406
<b>Hungary</b>	253582	<b>Denmark</b>	31454
<b>Norway</b>	235328	<b>Greece</b>	29611
<b>Georgia</b>	211795	<b>Lithuania</b>	28972
<b>Estonia</b>	194152	<b>Slovenia</b>	19947
<b>United Kingdom</b>	185773	<b>Croatia</b>	10997
<b>Latvia</b>	164870	<b>Luxembourg</b>	7223
<b>Slovakia</b>	157686	<b>Cyprus</b>	4598
<b>Netherlands</b>	108862	<b>Malta</b>	3569
<b>Italy</b>	105469	<b>Iceland</b>	1150
<b>Sweden</b>	69739	<b>Serbia</b>	93
<b>Ireland</b>	58967	<b>Armenia</b>	3

The following table contains links to data exports of cleaned publications describing public procurement in a particular jurisdiction. Each archive is encrypted by standard encryption software GnuPG. To decrypt archives on Windows, the installation of software like gpg4win (<https://www.gpg4win.org/download.html>) is required. Linux distributions and Mac operating systems usually have this software built-in. After the installation of gpg4win, the decrypt function is available in a context menu. Decryption also requires a password which we provide on demand by emailing [digiwhist.aws@gmail.com](mailto:digiwhist.aws@gmail.com).

<b>Poland</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/PL_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/PL_data.json.tar.gz.gpg</a>
<b>France</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/FR_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/FR_data.json.tar.gz.gpg</a>
<b>Portugal</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/PT_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/PT_data.json.tar.gz.gpg</a>
<b>Spain</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/ES_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/ES_data.json.tar.gz.gpg</a>
<b>Czech Republic</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/CZ_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/CZ_data.json.tar.gz.gpg</a>
<b>Germany</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/DE_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/DE_data.json.tar.gz.gpg</a>
<b>Hungary</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/HU_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/HU_data.json.tar.gz.gpg</a>
<b>Norway</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/NO_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/NO_data.json.tar.gz.gpg</a>
<b>Georgia</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/GE_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/GE_data.json.tar.gz.gpg</a>
<b>Estonia</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/EE_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/EE_data.json.tar.gz.gpg</a>

<b>United Kingdom</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/UK_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/UK_data.json.tar.gz.gpg</a>
<b>Latvia</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/LV_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/LV_data.json.tar.gz.gpg</a>
<b>Slovakia</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/SK_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/SK_data.json.tar.gz.gpg</a>
<b>Netherlands</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/NL_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/NL_data.json.tar.gz.gpg</a>
<b>Italy</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/IT_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/IT_data.json.tar.gz.gpg</a>
<b>Sweden</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/SE_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/SE_data.json.tar.gz.gpg</a>
<b>Ireland</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/IE_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/IE_data.json.tar.gz.gpg</a>
<b>Belgium</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/BE_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/BE_data.json.tar.gz.gpg</a>
<b>Romania</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/RO_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/RO_data.json.tar.gz.gpg</a>
<b>Bulgaria</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/BG_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/BG_data.json.tar.gz.gpg</a>
<b>Finland</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/FI_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/FI_data.json.tar.gz.gpg</a>
<b>Austria</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/AT_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/AT_data.json.tar.gz.gpg</a>
<b>Switzerland</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/CH_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/CH_data.json.tar.gz.gpg</a>
<b>Denmark</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/DK_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/DK_data.json.tar.gz.gpg</a>
<b>Greece</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/GR_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/GR_data.json.tar.gz.gpg</a>
<b>Lithuania</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/LT_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/LT_data.json.tar.gz.gpg</a>
<b>Slovenia</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/SI_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/SI_data.json.tar.gz.gpg</a>
<b>Croatia</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/HR_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/HR_data.json.tar.gz.gpg</a>
<b>Luxembourg</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/LU_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/LU_data.json.tar.gz.gpg</a>
<b>Cyprus</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/CY_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/CY_data.json.tar.gz.gpg</a>
<b>Malta</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/MT_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/MT_data.json.tar.gz.gpg</a>
<b>Iceland</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/IS_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/IS_data.json.tar.gz.gpg</a>
<b>Serbia</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/RS_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/RS_data.json.tar.gz.gpg</a>
<b>Armenia</b>	<a href="https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/AM_data.json.tar.gz.gpg">https://s3.eu-central-1.amazonaws.com/digiwhist-data/D2_5/tender/AM_data.json.tar.gz.gpg</a>

# References

Peter Lindes , Deryle W. Lonsdale , David W. Embley, Ontology-based information extraction with a cognitive agent, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, p.558-564, 2015

William B. Frakes, Ricardo Baeza-Yates, Information retrieval: data structures and algorithms, Prentice-Hall, Inc., Upper Saddle River, NJ, 1992

J. Hammer, H. Garcia-Molina, J. Cho, R. Aranha, and A. Crespo. Extracting semistructured information from the web. in Proceedings of the Workshop on Management of Semistructured Data, Tucson, Arizona, May 1997

David W. Embley , Mukkai S. Krishnamoorthy , George Nagy , Sharad Seth, Converting heterogeneous statistical tables on the web to searchable databases, International Journal on Document Analysis and Recognition, v.19 n.2, p.119-138, June 2016

Svátek V., Kompuš P., Dudáš M., Nečaský M., Klímek J.: Procurement notice enrichment using product ontologies, in Proceedings of the 11th International Conference on Semantic Systems, SEMANTICS 2015, Vienna, Austria, Association for Computing Machinery, ISBN: 978-1-4503-3462-4, pp. 200-203, 2015

# Annex 1

Information on data completeness for processed sources

## Data checks description

Tens of statistical values describing data quality and completeness are computed. Very often it's important to read these statistics globally not line by line. For example there are fields describing award criteria at both a tender and lot level because sources publish such data differently. This means that even if award criteria are missing at a tender level for 100% of contract notices this might not be a problem if award criteria are published at a lot level.

1st publication included	This check says whether the flag isIncluded is set for one of all publications. A publication marked as isIncluded = true contains data that can be found within a particular parsed/clean record
formType	Is the form type enumeration value set for a publication which is marked as isIncluded = true?
awardCriteria	Is at least one award criteria set at a tender level? This does not need to be necessarily an error. Award criteria can also be set at a lot level. This check is performed for contract notices only.
awardDeadline	Is award deadline set for a tender? This check is performed for contract notices only.
awardDecisionDate	Is award decision date set for a tender? This check is performed for contract awards only.
bidDeadline	Is bid deadline set for a tender? This check is performed for contract notices only.
buyer.address.city	Is city information set for an existing buyer's address?
buyer.address.country	Is country information set for an existing buyer's address?
buyer.address.rawAddress	Is an unstructured address set for an existing buyer's address? An unstructured address is set when address information cannot be split into structured information
buyer.address.street	Is street information set for an existing buyer's address?
buyer.bodyId	Is an ID variable of body identifier set for an existing body identifier at a buyer level? This checks whether

	the data does not contain body identifiers with only scope and type set which is useless.
buyer.bodyIds	Is at least one body identifier object set for an existing buyer?
buyer.buyerType	Is buyer type set for an existing buyer?
buyer.mainActivites	Is at least one main activity set for an existing buyer?
buyer.name	Is name set for an existing buyer?
buyers	Does at least one buyer exist for a tender?
country	Is the country of origin of a tender set?
cpvs	Is at least one CPV set at a tender level?
cpvs.code	When CPV exists at a tender level does it contain its code?
estimatedCompletionDate	Is an estimated completion date set for a tender? This check is performed for contract notices only.
estimatedStartDate	Is estimated start date set for a tender? This check is performed for contract notices only.
hasLots	Is information about whether a tender is divided into lots set for a tender?
isCoveredByGpa	Is tender covered by WTO Agreement on Government Procurement ( <a href="https://www.wto.org/english/tratop_e/gproc_e/gp_gp_a_e.htm">https://www.wto.org/english/tratop_e/gproc_e/gp_gp_a_e.htm</a> )
isDps	Is information about whether a dynamic purchasing system is used for a tender set?
isFrameworkAgreement	Is information about whether a tender is a framework agreement or not set?
lot.awardCriteria	Is at least one award criterion set for each existing lot? This check is performed for contract notices only.

lot.awardCriterion.name	When award criteria exist at a lot level do they have a name set?
lot.awardCriterion.weight	When award criteria exist at a lot level do they have a weight set?
lot.bid.bidder.address.city	Is city information set for an existing bidder's address?
lot.bid.bidder.address.country	Is country information set for an existing bidder's address?
lot.bid.bidder.address.rawAddress	Is an unstructured address set for an existing bidder's address? An unstructured address is set when address information cannot be split into structured information
lot.bid.bidder.address.street	Is street information set for an existing bidder's address?
lot.bid.bidder.bodyId.id	Is an ID variable of body identifier set for an existing body identifier at a bidder level? This checks whether the data does not contain body identifiers with only scope and type set which is useless.
lot.bid.bidder.bodyIds	Is at least one body identifier object set for an existing bidder?
lot.bid.bidder.name	Is a name set for an existing bidder?
lot.bid.bidders	Does at least one bidder exist for an existing bid
lot.bid.price	Is a price set for an existing bid? This check is performed for contract awards only.
lot.bids	Does at least one bid exist for an existing lot? This check is performed for contract awards only.
lot.contractNumber	Is contract number set for an existing lot?
lot.cpv.code	When CPV exists at a lot level does it contain its code?
lot.cpv.s	Is at least one CPV set for an existing lot?
lot.estimatedPrice	Is estimated price set for an existing lot?

lot.isAwarded	Is information about whether a lot has already been awarded or not set for an existing lot?
lot.lotNumber	Is a lot number set for a lot?
lot.notCancelledWinninBid	Does at least one non-cancelled lot with a winning bid exist for a contract award?
lot.positionOnPage	Is information about position on page set for an existing lot?
lot.status	Is status set for an existing lot?
lot.title	Is title set for an existing lot?
lots	Does contract award contain at least one lot?
nationalProcedureType	Is national procedure type set for a tender?
npwpReasons	Are reasons for a negotiated procedure without publication set when a tender procedure type is set to negotiated procedure without publication?
procedureType	Is enumeration value for a procedure type set for a publication?
publication.formType	Is form type enumeration value set for an existing publication?
publication.publicationDate	Is a publication date set for an existing publication?
publication.publicationURL	Is a human or machine readable url date set for an existing publication?
publication.sourceFormType	Is source form type set for an existing publication?
publications	Is at least one publication set for a tender?
publications.isIncluded	Is a flag which marks whether data from a linked publication is included?

---

selectionMethod

Is selection method set for a tender?

---

supplyType

Is supply type set for a tender?

---

tender.estimatedPrice

Is estimated price set for a tender?

---

tender.finalPrice

Is final price set for a tender?

---

title

Is title set for a tender?

---

buyerAssignedId

Is a buyer assigned ID set for a tender?

---

## Tenders Electronic Daily

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriteria	55.078	44.922
awardCriterion.name	100	0
awardCriterion.weight	95.483	4.517
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	100	0
buyer.address.city	100	0
buyer.address.country	100	0
buyer.address.rawAddress	0	100
buyer.address.street	97.752	2.248
buyer.bodyId	100	0
buyer.bodyIds	15.543	84.457
buyer.buyerType	82.502	17.498
buyer.mainActivites	0	100
buyer.name	100	0
buyers	100	0
country	100	0
cpvs	100	0
cpvs.code	100	0
estimatedCompletionDate	19.531	80.469
estimatedStartDate	19.922	80.078

hasLots	56.6	43.4
isCoveredByGpa	100	0
isDps	100	0
isFrameworkAgreement	100	0
lot.awardCriteria	10.477	89.523
lot.awardCriterion.name	100	0
lot.awardCriterion.weight	64.871	35.129
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	100	0
lot.bid.bidder.bodyIds	11.193	88.807
lot.bid.bidder.name	100	0
lot.bid.bidders	91.976	8.024
lot.bid.price	66.238	33.762
lot.bids	98.561	1.439
lot.contractNumber	32.913	67.087
lot.cpv.code	100	0
lot.cpv.s	55.617	44.383
lot.estimatedPrice	32.598	67.402
lot.isAwarded	0	100
lot.lotNumber	80.41	19.59
lot.notCancelledWinninBid	100	0
lot.positionOnPage	100	0

lot.status	99.724	0.276
lot.title	89.87	10.13
lots	98.973	1.027
nationalProcedureType	0	100
procedureType	99.4	0.6
publication.formType	72.512	27.488
publication.publicationDate	95.122	4.878
publication.publicationURL	100	0
publication.sourceFormType	72.512	27.488
publications	100	0
publications.isIncluded	100	0
selectionMethod	91.9	8.1
supplyType	100	0
tender.estimatedPrice	15.7	84.3
tender.finalPrice	58.111	41.889
title	97.7	2.3
buyerAssignedId	0	100

# Spain

<b>Check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriteria	0	100
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	99.251	0.749
buyer.address.city	100	0
buyer.address.country	100	0
buyer.address.rawAddress	0	100
buyer.address.street	100	0
buyer.bodyId	100	0
buyer.bodyIds	100	0
buyer.buyerType	0	100
buyer.mainActivites	0	100
buyer.name	100	0
buyers	74	26
country	100	0
cpvs	74	26
cpvs.code	98.297	1.703
estimatedCompletionDate	0	100
estimatedStartDate	0	100
hasLots	0	100
isCoveredByGpa	0	100

isDps	0	100
isFrameworkAgreement	0	100
lot.awardCriteria	0	100
lot.awardCriterion.name	0	100
lot.awardCriterion.weight	0	100
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	97.12	2.88
lot.bid.bidder.bodyIds	100	0
lot.bid.bidder.name	97.12	2.88
lot.bid.bidders	100	0
lot.bid.price	89.058	10.942
lot.bids	52.429	47.571
lot.contractNumber	0	100
lot.cpv.code	99.721	0.279
lot.cpv.s	60.05	39.95
lot.estimatedPrice	55.616	44.384
lot.isAwarded	0	100
lot.lotNumber	60.05	39.95
lot.notCancelledWinninBid	100	0
lot.positionOnPage	0	100
lot.status	94.102	5.898
lot.title	60.05	39.95

lots	100	0
nationalProcedureType	0	100
npwpReasons	0	100
procedureType	71.98	28.02
publication.formType	54.413	45.587
publication.publicationDate	40.266	59.734
publication.publicationURL	100	0
publication.sourceFormType	54.413	45.587
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100
supplyType	74	26
tender.estimatedPrice	69.38	30.62
tender.finalPrice	0	100
title	73.1	26.9
buyerAssignedId	73.1	26.9

## Estonia

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriteria	81.857	18.143
awardCriterion.name	100	0
awardCriterion.weight	100	0
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	0	100
buyer.address.city	99.798	0.202
buyer.address.country	100	0
buyer.address.rawAddress	0	100
buyer.address.street	100	0
buyer.bodyId	100	0
buyer.bodyIds	100	0
buyer.buyerType	86.332	13.668
buyer.mainActivites	0	100
buyer.name	100	0
buyers	69.36	30.64
country	100	0
cpvs	68.6	31.4
cpvs.code	100	0
estimatedCompletionDate	0	100
estimatedStartDate	0	100

hasLots	27.9	72.1
isCoveredByGpa	68.56	31.44
isDps	27.9	72.1
isFrameworkAgreement	68.56	31.44
lot.awardCriteria	0	100
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	100	0
lot.bid.bidder.bodyIds	100	0
lot.bid.bidder.name	100	0
lot.bid.bidders	0.131	99.869
lot.bid.price	100	0
lot.bids	100	0
lot.contractNumber	0	100
lot.cpv.code	100	0
lot.cpv.s	9.877	90.123
lot.estimatedPrice	3.704	96.296
lot.isAwarded	0	100
lot.lotNumber	59.73	40.27
lot.notCancelledWinninBid	100	0
lot.positionOnPage	0	100
lot.status	10.053	89.947
lot.title	100	0

lots	0.098	99.902
nationalProcedureType	68.56	31.44
procedureType	0	100
publication.formType	91.274	8.726
publication.publicationDate	28.356	71.644
publication.publicationURL	100	0
publication.sourceFormType	91.274	8.726
publications	100	0
publications.isIncluded	100	0
selectionMethod	66.74	33.26
supplyType	0	100
tender.estimatedPrice	2.32	97.68
tender.finalPrice	94.201	5.799
title	100	0
buyerAssignedId	0	100

# Hungary

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	0	100
buyer.address.city	3.3	96.7
buyer.address.country	3.3	96.7
buyer.address.rawAddress	0	100
buyer.address.street	3.3	96.7
buyer.bodyId	3.1	96.9
buyer.bodyIds	100	0
buyer.buyerType	1.95	98.05
buyer.mainActivites	0	100
buyer.name	3.15	96.85
buyers	100	0
country	100	0
cpvs	100	0
cpvs.code	97.75	2.25
hasLots	0	100
isCoveredByGpa	0	100
isDps	0	100
isFrameworkAgreement	0	100
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100

lot.bid.bidder.bodyId.id	15.116	84.884
lot.bid.bidder.bodyIds	100	0
lot.bid.bidder.name	31.395	68.605
lot.bid.bidders	100	0
lot.contractNumber	0	100
lot.cpv5	0	100
lot.estimatedPrice	11.628	88.372
lot.isAwarded	0	100
lot.lotNumber	37.209	62.791
lot.positionOnPage	100	0
lot.status	0	100
lot.title	37.209	62.791
nationalProcedureType	0	100
procedureType	0	100
publication.formType	0	100
publication.publicationDate	0	100
publication.publicationURL	100	0
publication.sourceFormType	100	0
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100
supplyType	73.65	26.35
tender.estimatedPrice	0	100
title	76.65	23.35
buyerAssignedId	0	100

# Slovakia

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	92.5	7.5
awardCriteria	88.312	11.688
awardCriterion.name	100	0
awardCriterion.weight	84.868	15.132
awardDeadline	48.052	51.948
awardDecisionDate	0	100
bidDeadline	36.364	63.636
buyer.address.city	5.5	94.5
buyer.address.country	65.5	34.5
buyer.address.rawAddress	83	17
buyer.address.street	5.5	94.5
buyer.bodyId	88.5	11.5
buyer.bodyIds	100	0
buyer.buyerType	74	26
buyer.mainActivites	0	100
buyer.name	88.5	11.5
buyers	100	0
cancellationDate	100	0
contractCancellation	0	100
country	100	0
cpvs	80.5	19.5
cpvs.code	100	0

estimatedCompletionDate	0	100
estimatedStartDate	0	100
hasLots	32.5	67.5
isCoveredByGpa	30	70
isDps	0	100
isFrameworkAgreement	29	71
lot.awardCriteria	0	100
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	100	0
lot.bid.bidder.bodyIds	100	0
lot.bid.bidder.name	100	0
lot.bid.bidders	100	0
lot.bid.price	98.333	1.667
lot.bids	98.276	1.724
lot.contractNumber	0	100
lot.cpv.code	82.609	17.391
lot.cpv.s	16.912	83.088
lot.estimatedPrice	62.5	37.5
lot.isAwarded	14.706	85.294
lot.lotNumber	60.294	39.706
lot.notCancelledWinninBid	100	0
lot.positionOnPage	100	0

lot.status	58.824	41.176
lot.title	82.353	17.647
lots	100	0
nationalProcedureType	38.5	61.5
procedureType	38.5	61.5
publication.formType	93.678	6.322
publication.publicationDate	22.989	77.011
publication.publicationURL	100	0
publication.sourceFormType	93.678	6.322
publications	100	0
publications.isIncluded	100	0
selectionMethod	59.5	40.5
supplyType	57	43
tender.estimatedPrice	59.5	40.5
tender.finalPrice	95	5
title	80.5	19.5
buyerAssignedId	0	100

## Czech Republic

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriteria	80.247	19.753
awardCriterion.name	100	0
awardCriterion.weight	98.876	1.124
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	100	0
buyer.address.city	100	0
buyer.address.country	100	0
buyer.address.rawAddress	0	100
buyer.address.street	100	0
buyer.bodyId	100	0
buyer.bodyIds	71.25	28.75
buyer.buyerType	88.75	11.25
buyer.mainActivites	0	100
buyer.name	100	0
buyers	80	20
cancellationDate	0	100
contractCancellation	0	100
country	100	0
cpvs	78	22
cpvs.code	100	0

estimatedCompletionDate	37.037	62.963
estimatedStartDate	32.099	67.901
hasLots	3.5	96.5
isCoveredByGpa	66	34
isDps	4	96
isFrameworkAgreement	68	32
lot.awardCriteria	13.889	86.111
lot.awardCriterion.name	100	0
lot.awardCriterion.weight	100	0
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	100	0
lot.bid.bidder.bodyIds	98.333	1.667
lot.bid.bidder.name	99.444	0.556
lot.bid.bidders	100	0
lot.bid.price	98.256	1.744
lot.bids	100	0
lot.contractNumber	53.081	46.919
lot.cpv.code	100	0
lot.cpv.s	18.009	81.991
lot.estimatedPrice	81.991	18.009
lot.isAwarded	1.896	98.104
lot.lotNumber	74.408	25.592

lot.notCancelledWinninBid	100	0
lot.positionOnPage	99.526	0.474
lot.status	97.63	2.37
lot.title	91.943	8.057
lots	96.667	3.333
nationalProcedureType	77	23
procedureType	78.5	21.5
publication.formType	85.532	14.468
publication.publicationDate	93.333	6.667
publication.publicationURL	100	0
publication.sourceFormType	85.532	14.468
publications	100	0
publications.isIncluded	100	0
selectionMethod	62	38
supplyType	77	23
tender.estimatedPrice	34	66
tender.finalPrice	95	5
title	79.5	20.5
buyerAssignedId	0	100

# Latvia

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriteria	17.44	82.56
awardCriterion.name	100	0
awardCriterion.weight	100	0
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	21.205	78.795
buyer.address.city	83.916	16.084
buyer.address.country	97.752	2.248
buyer.address.rawAddress	0	100
buyer.address.street	97.752	2.248
buyer.bodyId	100	0
buyer.bodyIds	97.752	2.248
buyer.buyerType	0	100
buyer.mainActivites	0	100
buyer.name	97.752	2.248
buyers	100	0
country	100	0
cpvs	83.916	16.084
cpvs.code	39.074	60.926
estimatedCompletionDate	3.764	96.236
estimatedStartDate	2.008	97.992

hasLots	100	0
isCoveredByGpa	100	0
isDps	100	0
isFrameworkAgreement	100	0
lot.awardCriteria	0	100
lot.bids	0	100
lot.contractNumber	0	100
lot.cpv.code	28.218	71.782
lot.cpv.s	100	0
lot.estimatedPrice	7.019	92.981
lot.isAwarded	0	100
lot.lotNumber	98.98	1.02
lot.positionOnPage	0	100
lot.status	64.067	35.933
lot.title	99.4	0.6
lots	24.427	75.573
nationalProcedureType	8.442	91.558
procedureType	8.442	91.558
publication.formType	68.819	31.181
publication.publicationDate	69.643	30.357
publication.publicationURL	100	0
publication.sourceFormType	68.819	31.181
publications	100	0
publications.isIncluded	100	0
selectionMethod	19.031	80.969

supplyType	37.562	62.438
tender.estimatedPrice	0	100
tender.finalPrice	0	100
title	83.916	16.084
buyerAssignedId	83.766	16.234

## Norway

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	0	100
buyer.address.city	3.3	96.7
buyer.address.country	3.3	96.7
buyer.address.rawAddress	0	100
buyer.address.street	3.3	96.7
buyer.bodyId	3.1	96.9
buyer.bodyIds	100	0
buyer.buyerType	1.95	98.05
buyer.mainActivites	0	100
buyer.name	3.15	96.85
buyers	100	0
country	100	0
cpvs	100	0
cpvs.code	97.75	2.25
hasLots	0	100
isCoveredByGpa	0	100
isDps	0	100
isFrameworkAgreement	0	100
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100

lot.bid.bidder.bodyId.id	15.116	84.884
lot.bid.bidder.bodyIds	100	0
lot.bid.bidder.name	31.395	68.605
lot.bid.bidders	100	0
lot.contractNumber	0	100
lot.cpv5	0	100
lot.estimatedPrice	11.628	88.372
lot.isAwarded	0	100
lot.lotNumber	37.209	62.791
lot.positionOnPage	100	0
lot.status	0	100
lot.title	37.209	62.791
nationalProcedureType	0	100
procedureType	0	100
publication.formType	0	100
publication.publicationDate	0	100
publication.publicationURL	100	0
publication.sourceFormType	100	0
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100
supplyType	73.65	26.35
tender.estimatedPrice	0	100
title	76.65	23.35
buyerAssignedId	0	100

# France

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriterion.name	96.167	3.833
awardCriterion.weight	95.006	4.994
buyer.address.city	99.7	0.3
buyer.address.country	0.15	99.85
buyer.address.rawAddress	13.25	86.75
buyer.address.street	86.3	13.7
buyer.bodyIds	0	100
buyer.buyerType	0	100
buyer.mainActivites	0	100
buyer.name	86.3	13.7
buyers	100	0
cancellationDate	0	100
country	100	0
cpvs	35.1	64.9
cpvs.code	99.616	0.384
hasLots	13.4	86.6
isCoveredByGpa	0	100
isDps	13.4	86.6
isFrameworkAgreement	13.4	86.6
lot.awardCriterion.name	0	100
lot.awardCriterion.weight	0	100

lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	1.596	98.404
lot.bid.bidder.bodyIds	100	0
lot.bid.bidder.name	100	0
lot.bid.bidders	43.069	56.931
lot.contractNumber	10.572	89.428
lot.cpv5	0	100
lot.estimatedPrice	0	100
lot.isAwarded	0	100
lot.lotNumber	22.977	77.023
lot.positionOnPage	0	100
lot.status	0	100
lot.title	72.6	27.4
nationalProcedureType	0	100
procedureType	0	100
publication.formType	100	0
publication.publicationDate	82	18
publication.publicationURL	100	0
publication.sourceFormType	100	0
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100

supplyType	0	100
tender.estimatedPrice	0.15	99.85
title	64.8	35.2
buyerAssignedId	0	100

# Poland

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriteria	0	100
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	100	0
buyer.address.city	100	0
buyer.address.country	0	100
buyer.address.rawAddress	99.3	0.7
buyer.address.street	99.3	0.7
buyer.bodyId	94.2	5.8
buyer.bodyIds	100	0
buyer.buyerType	88.8	11.2
buyer.mainActivites	0	100
buyer.name	100	0
buyers	100	0
country	100	0
cpvs	88.8	11.2
cpvs.code	100	0
estimatedCompletionDate	0	100
estimatedStartDate	0	100
hasLots	0	100
isCoveredByGpa	0	100

isDps	0	100
isFrameworkAgreement	0	100
lot.awardCriteria	0	100
lot.awardCriterion.name	100	0
lot.awardCriterion.weight	13.636	86.364
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyIds	0	100
lot.bid.bidder.name	100	0
lot.bid.bidders	100	0
lot.bids	0	100
lot.contractNumber	0	100
lot.cpv.code	100	0
lot.cpv.s	29.817	70.183
lot.estimatedPrice	59.513	40.487
lot.isAwarded	0	100
lot.lotNumber	72.292	27.708
lot.positionOnPage	0	100
lot.status	100	0
lot.title	70.953	29.047
lots	17.613	82.387
nationalProcedureType	0	100
npwpReasons	0	100

procedureType	88.8	11.2
publication.formType	48.379	51.621
publication.publicationDate	81.219	18.781
publication.publicationURL	100	0
publication.sourceFormType	48.379	51.621
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100
supplyType	88.8	11.2
tender.estimatedPrice	44.25	55.75
tender.finalPrice	0	100
title	88.8	11.2
buyerAssignedId	0	100

# Ireland

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriteria	0	100
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	0	100
buyer.address.city	89.837	10.163
buyer.address.country	89.837	10.163
buyer.address.rawAddress	0	100
buyer.address.street	0	100
buyer.bodyIds	0	100
buyer.buyerType	0	100
buyer.mainActivites	0	100
buyer.name	89.837	10.163
buyers	100	0
country	100	0
cpvs	43.598	56.402
cpvs.code	100	0
estimatedCompletionDate	0	100
estimatedStartDate	0	100
hasLots	0	100
isCoveredByGpa	0	100
isDps	0	100

isFrameworkAgreement	0	100
lots	0	100
nationalProcedureType	0	100
procedureType	0	100
publication.formType	100	0
publication.publicationDate	0	100
publication.publicationURL	0	100
publication.sourceFormType	100	0
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100
supplyType	98.069	1.931
tender.estimatedPrice	0	100
tender.finalPrice	0	100
title	100	0
buyerAssignedId	0	100

# Netherlands

Check	ok [%]	error [%]
1st publication included	100	0
formType	100	0
awardCriteria	8.151	91.849
awardCriterion.name	100	0
awardCriterion.weight	0	100
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	0	100
buyer.address.city	50.193	49.807
buyer.address.country	50.193	49.807
buyer.address.rawAddress	43.52	56.48
buyer.address.street	50.193	49.807
buyer.bodyId	100	0
buyer.bodyIds	93.713	6.287
buyer.buyerType	48.725	51.275
buyer.mainActivites	0	100
buyer.name	93.713	6.287
buyers	72.74	27.26
country	100	0
cpvs	72.74	27.26
cpvs.code	100	0
estimatedCompletionDate	33.042	66.958
estimatedStartDate	33.382	66.618

hasLots	23.02	76.98
isCoveredByGpa	33.24	66.76
isDps	18	82
isFrameworkAgreement	58.32	41.68
lot.awardCriteria	0	100
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	35.422	64.578
lot.bid.bidder.bodyIds	9.831	90.169
lot.bid.bidder.name	98.393	1.607
lot.bid.bidders	99.946	0.054
lot.bid.price	12.905	87.095
lot.bids	94.622	5.378
lot.contractNumber	7.028	92.972
lot.cpv.code	100	0
lot.cpv.s	20.919	79.081
lot.estimatedPrice	6.517	93.483
lot.isAwarded	0	100
lot.lotNumber	68.927	31.073
lot.notCancelledWinninBid	100	0
lot.positionOnPage	0	100
lot.status	100	0
lot.title	45.822	54.178

lots	91.623	8.377
nationalProcedureType	72.74	27.26
procedureType	0	100
publication.formType	96.615	3.385
publication.publicationDate	96.147	3.853
publication.publicationURL	100	0
publication.sourceFormType	96.615	3.385
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100
supplyType	33.78	66.22
tender.estimatedPrice	0	100
tender.finalPrice	0	100
title	72.74	27.26
buyerAssignedId	26.1	73.9

# Georgia

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	100	0
awardCriteria	0	100
awardDeadline	0	100
awardDecisionDate	0	100
bidDeadline	100	0
buyer.address.city	100	0
buyer.address.country	100	0
buyer.address.rawAddress	0	100
buyer.address.street	100	0
buyer.bodyId	100	0
buyer.bodyIds	100	0
buyer.buyerType	100	0
buyer.mainActivites	0	100
buyer.name	100	0
buyers	100	0
contractCancellation	0	100
country	100	0
cpvs	100	0
cpvs.code	100	0
estimatedCompletionDate	0	100
estimatedStartDate	0	100
hasLots	0	100

isCoveredByGpa	0	100
isDps	0	100
isFrameworkAgreement	0	100
lot.awardCriteria	0	100
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	100	0
lot.bid.bidder.bodyIds	100	0
lot.bid.bidder.name	99.861	0.139
lot.bid.bidders	100	0
lot.bid.price	100	0
lot.bids	99.925	0.075
lot.contractNumber	0	100
lot.cpvs	0	100
lot.estimatedPrice	0	100
lot.isAwarded	0	100
lot.lotNumber	0	100
lot.notCancelledWinninBid	100	0
lot.positionOnPage	0	100
lot.status	99.15	0.85
lot.title	0	100
lots	100	0
nationalProcedureType	0	100

procedureType	100	0
publication.formType	100	0
publication.publicationDate	100	0
publication.publicationURL	100	0
publication.sourceFormType	100	0
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100
supplyType	0	100
tender.estimatedPrice	98.05	1.95
tender.finalPrice	0	100
title	0	100
buyerAssignedId	0	100

## United Kingdom

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	43.275	56.725
awardDecisionDate	0	100
buyer.address.city	70.428	29.572
buyer.address.country	56.226	43.774
buyer.address.rawAddress	0	100
buyer.address.street	71.595	28.405
buyer.bodyId	100	0
buyer.bodyIds	43.191	56.809
buyer.buyerType	43.191	56.809
buyer.mainActivites	0	100
buyer.name	82.49	17.51
buyers	100	0
country	43.191	56.809
cpvs	99.805	0.195
cpvs.code	100	0
hasLots	0	100
isCoveredByGpa	0	100
isDps	0	100
isFrameworkAgreement	56.809	43.191
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100

lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyIds	0	100
lot.bid.bidder.name	58.076	41.924
lot.bid.bidders	56.725	43.275
lot.bid.price	0	100
lot.bids	100	0
lot.contractNumber	0	100
lot.cpvs	0	100
lot.estimatedPrice	0	100
lot.isAwarded	0	100
lot.lotNumber	0	100
lot.positionOnPage	0	100
lot.status	59.728	40.272
lot.title	0	100
lots	100	0
nationalProcedureType	49.903	50.097
procedureType	0	100
publication.formType	19.681	80.319
publication.publicationDate	45.567	54.433
publication.publicationURL	100	0
publication.sourceFormType	45.567	54.433
publications	100	0
publications.isIncluded	100	0
selectionMethod	0	100
supplyType	67.251	32.749

tender.estimatedPrice	0	100
tender.finalPrice	0	100
title	100	0
buyerAssignedId	42.105	57.895

# Portugal

<b>check</b>	<b>ok [%]</b>	<b>error [%]</b>
1st publication included	100	0
formType	6.526	93.474
awardCriteria	0	100
awardDeadline	0	100
bidDeadline	0	100
buyer.address.city	0	100
buyer.address.country	0	100
buyer.address.rawAddress	0	100
buyer.address.street	0	100
buyer.bodyId	100	0
buyer.bodyIds	100	0
buyer.buyerType	0	100
buyer.mainActivites	0	100
buyer.name	100	0
buyers	100	0
country	100	0
cpvs	92.131	7.869
cpvs.code	100	0
estimatedCompletionDate	0	100
estimatedStartDate	0	100
hasLots	0	100
isCoveredByGpa	0	100
isDps	0	100

isFrameworkAgreement	0	100
lot.bid.bidder.address.city	0	100
lot.bid.bidder.address.country	0	100
lot.bid.bidder.address.rawAddress	0	100
lot.bid.bidder.address.street	0	100
lot.bid.bidder.bodyId.id	100	0
lot.bid.bidder.bodyIds	99.44	0.56
lot.bid.bidder.name	99.44	0.56
lot.bid.bidders	100	0
lot.contractNumber	0	100
lot.cpv	0	100
lot.estimatedPrice	0	100
lot.isAwarded	0	100
lot.lotNumber	0	100
lot.positionOnPage	0	100
lot.status	0	100
lot.title	0	100
nationalProcedureType	100	0
procedureType	0	100
publication.formType	5.511	94.489
publication.publicationDate	87.682	12.318
publication.publicationURL	100	0
publication.sourceFormType	5.511	94.489
publications	100	0
publications.isIncluded	100	0

selectionMethod	0	100
supplyType	3.839	96.161
tender.estimatedPrice	5.758	94.242
title	0	100
buyerAssignedId	0	100