



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: D2.3 Data Template for information on public procurement tendering and its actors

Due date of deliverable: 31/10/2015

Actual submission date: 27/10/2015

Authors: Ágnes Czibik; Bence Tóth; István János Tóth

Organization name of lead beneficiary for this deliverable: Átlátható Kormányzás Kutatóintézet Kft.

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

Ágnes Czibik Átlátható Kormányzás Kutatóintézet
Bence Tóth Cambridge University
István János Tóth Átlátható Kormányzás Kutatóintézet

27 October 2015

ABSTRACT – This report takes a step further and translates the results of the comprehensive legal mapping of information on public procurement tendering into a data template. This template may be considered a pragmatic and detailed documentation of the dataset that DIGIWHIST aims to retrieve. Part I outlines the database structure that provides DIGIWHIST with a robust structure which incorporates public procurement information in 35 European jurisdictions as well as information on companies, public sector bodies, and political office holders. Part II defines the variables; specifies their data types and their position and role within the data structure. Part III presents the availability of public procurement data in different countries in terms of exactly defined variables and Part IV provides a review of the connection between DIGIWHIST and the Open Contracting Data Standard emphasising the importance of transparent, standard publication of good quality data on public procurement.

KEYWORDS: public procurement, database structure, data template

Corresponding author: Ágnes Czibik
Email: aczibik@govtransparency.eu
Átlátható Kormányzás Kutatóintézet

////////////////////////////////////
© 2015 Átlátható Kormányzás Kutatóintézet

All rights reserved. This project has received funding from the European Union’s Horizon 2020 research and innovation Programme under grant agreement No 645852.

The information and views set out in this publication are those of the author(s) only and do not reflect any collective opinion of the DIGIWHIST consortium, nor do they reflect the official opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the European Commission is responsible for the use which might be made of the following information.

Contents

I. PART. General Overview of the Structure of the Databases	5
1. The concept of the database structure	6
2. Variable types and formats	8
II. PART. Description of variables	9
1. Public procurement tender data	10
'Contract' object	10
'Lot' object	16
'Bid' object	19
'Winning Bid' object (Extends 'Bid')	20
'Publication' object	21
'Body' object	22
'Body Identifier' object	22
'Buyer' object	23
'Bidder' object	24
'Funding' object	24
'Award Criterion' object	25
'Qualification' object	25
'Payment' object	25
'Document' object	26
'Agreement' object	26
'Price' object	27
'Address' object	28
2. Company data	29
Registry data	29
Financial data	31
Ownership data	34
Managers/directors data	36
3. Public authority data	38
Budget data	39
4. Public officials	41
III. PART. Availability Of Public Procurement Data	42
IV. PART. DIGIWHIST and Open Contracting Data Standard comparison	46

List of Tables

Table 1: Data availability in published tender documents across countries	42
Table 2: <i>Data availability in published tender documents across countries (part II)</i>	43
Table 3: <i>Data availability in published tender documents across countries (part III)</i>	44
Table 4: <i>Data availability in published tender documents across countries (part IV)</i>	45
Table 5: Comparison of DIGIWHIST and Open Contracting Data Standard variables.....	46
Table 6: Variables that are available in OCDS's standard but not in DIGIWHIST	53

List of figures

Figure 1: Outline of the database structure related to contract award notices	7
---	---

I. PART. GENERAL OVERVIEW OF THE STRUCTURE OF THE DATABASES

1. The concept of the database structure

DIGIWHIST uses four main types of data sources; the most complex and fundamental information derives from (1) public procurement tenders. Beside this we use (2) company data, (3) data on public organisations and state-owned companies, including financial information, if available; and (4) information about public officials.

The challenge of database building resides in the substantive and formal diversity of data coming from different sources. We aim at proposing a robust standard which is able to incorporate all necessary data types in a flexible and reasonable system. Our approach is to follow 'linked open data principles'¹, which describe a method based on Web technologies for publishing structured data. It builds on interlinking data in order to enable computers to read automatically this data. The method makes it easier to connect and query data from different sources.

We implemented these principles by using and extending the Open Contracting Data Standard². The emerging database structure reasonably combines an ideal structure with real available data. According to the planned schedule in the Description of Work, we are in the middle of data collection which necessarily implies that this data template is only a guidance to be updated and finalised as we complete our data collection tasks. In short, the template represents the maximum possible or desirable database structure which is subject to, for example, actual data gaps, non-machine readable data formats or changing public data provision standards at the source. As the publicly available data have been changing dynamically and continue to do so, we aimed at finding the smallest common denominator allowing for consistent comparisons over time within the same country, but also cross-sectionally across countries and regions.

As public procurement data is the most complex, we use it as an example to present the concept of the database. Our standard is structured according to nested objects, which means broadly that variables are arranged in objects ('variable sets') and these objects are in hierarchical relationship with each other. In public procurement data the main unit – the starting point - is a contract. A contract is characterised on the one hand by single variables like official identifiers, title, description and procedure type. On the other hand it may be described by objects such as the 'buyer' object which specifies the contracting authority or other purchasing body of the tender (the buyer); but not with a single variable but with a variable set instead. This variable set consists of the identifier number, name, address, the main activity etc. of the purchasing body. If we go further, 'address' is not a single variable either, but it is an object containing variables like city, street, postcode, country. In this way the objects 'contract', 'buyer' and 'address' are nested.

Some specific objects - such as 'address' in the previous example - are used in connection with many other objects, in this specific case everywhere where the address of a public organisation or company is mentioned.

In some cases more variables or objects of the same kind are connected to the parent object, e.g. a company has more than one 'official identifier' (e.g. VAT number and registry number) and we would like to store both.

¹ https://en.wikipedia.org/wiki/Linked_data

² <http://standard.open-contracting.org/>

In Figure 1 we present the concept of the database structure with a more pragmatic example. As the figure shows, the 'contract' object has a central role in the structure. It has some characteristics that are stored in single variables such as the type of procedure. These variables can be string or numeric variables; the exact types we use are specified later in this chapter.

Beside simple variables, connected variable sets can describe an attribute of contracts such as the issuer of the tender, in other words the buyer. The identity of the buyer is an attribute of the contract, but this identity is described with a set of variables e.g. the name of the buyer, the address of it, its main activity. This variable set is referred to as an 'object'.

Eventually, a nested, multidimensional structure emerges from these objects. For example 'lot' object specifies the contract parts (lots) that the contract consists of; 'bid' object specifies a bid that was received for a lot; and 'bidder' object describes a bidder company which submitted a bid for a lot. In this way 'bidder' is an attribute of 'bid'; 'bid' is an attribute of 'lot'; and 'lot' is an attribute of 'contract'.

In some cases more attributes of the same kind are linked to the parent object, e.g. a contract may have more lots or a lot may have more bids.

Figure 1: Outline of database structure related to contract award notices

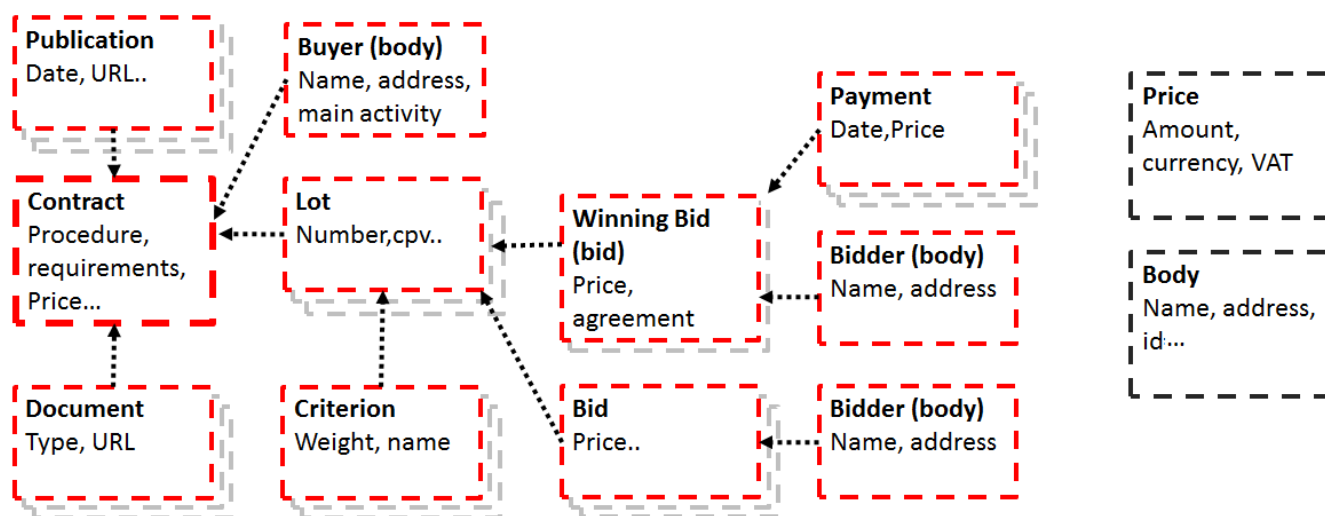


Figure 1 presents only a general overview on the database structure, in reality there are more objects and a more complex structure of connections between them. In Part II we define these objects and describe their content in detail. The full documentation is also available online at the Wikipedia of the project here: <http://digiwhist.datlab.cz/>. This report and the corresponding online Wikipedia page will form the core elements of the full documentation to be released along with the data.

2. Variable types and formats

We use the following data types and formats in Part II.

- *String* variables are character variables. They may contain numeric values too, but if they are defined as string, they cannot be used to do arithmetic operations with. They are typically used to store words and text e.g. names, addresses, unstructured text about the award criteria of a tender.
- *Enumerated* variables are categorical variables in statistics. They consist of a set of named values called elements, members or enumerators of the type. We use this type, if the entire set of possible values is well known, such as the status of the contract lot (prepared, announced, awarded, cancelled, finished) or the type of the subject of the contract (supplies, services, works)
- *Boolean* variables have two values (usually denoted *true* and *false*).
- *Integer* variables take an integer value (whole numbers).
- *Float and double* type variables both represent real values but they have different pre-defined limits on both their maximum values and their precision. They are often represented as decimal numbers.
- *Arrays* are used to store multiple values in a single variable for example more bidders connected to one lot. 'Value' may mean either a single variable or an object.

II. PART. DESCRIPTION OF VARIABLES

1. Public procurement tender data

‘Contract’ object

This is the main object in the structure; it contains most information on a single public procurement contract.

- id (int)

A unique identification number will be assigned to each contract. This is an integer variable. It contains positive whole numbers.

- URI (URL)

A unique URI (Uniform Resource Identifier) will be assigned to each contract. This variable contains values in URL format.

- announcementType (string)

This variable shows the type of the announcement where the information originates from. It is in string format.

- country (enum)

The ‘country’ variable contains information about the source country of the contract. It is an enumerated variable with values based on ‘ISO 3166-1 alpha-2’ codes which are standard two-letter codes representing countries, dependent territories, and special areas of geographical interest. More information: http://www.iso.org/iso/country_codes

- assignedContractIds (array - string)

This variable contains reference numbers given to contracts by buyer (public authority, state owned company etc.), the system of the national public procurement authority or external system (such as TED). This identifier helps establishing a connection between the data in DIGIWHIST’s databases and original sources. As data formats could vary in original sources (numbers, letters, special characters, different length) this variable is stored as a string in the DIGIWHIST database. This is an array: one contract may have more reference numbers assigned to it.

- assignedAnnouncementIds (array - string)

This variable contains reference numbers given to public procurement announcements by buyer, national public procurement system, or external system (such as TED). It is a string array: more announcement IDs may be assigned to one contract.

- relatedAnnouncementIds (array - string)

‘relatedAnnouncementIds’ is a string array containing reference numbers given to related announcements which are mentioned in the given announcement. The reference number is given by either buyer, system, or external system (such as TED).

- title (string)

This variable shows the title of the contract in original language in string format.

- titleEnglish (string)

The variable shows the title of the contract in English in string format.

- procedureType (enum)

The 'procedureType' variable identifies the type of public procurement procedures. If the national procedure types do not fit into DIGIWHIST's pre-defined typology, the national procedure type is stored in the 'nationalProcedureType' variable in string format.

It is an enumerated variable with the following values:

open, restricted, negotiated with publication, negotiated without publication, competitive dialog, design contest, framework agreement based contract, dynamic purchasing system based contract, direct award to single body, sealed bid competition between several bodies, sealed bid competition with open call, negotiated competition with open call.

- nationalProcedureType (string)

Type of procedure in the national databases – unstructured, string format.

- description (string)

This variable contains the description of the subject of the contract in original language in string format.

- descriptionEnglish (string)

This variable contains the description of the subject of the contract in English in string format.

- bidderLimitation (int)

Maximum number of bids, typically used in 'Restricted' or 'Negotiated with publication' procedures. Integer variable, positive whole numbers.

- estimatedValue (object – 'price')

'EstimatedValue' is an object containing information about the estimated value of the contract using the variables in 'Price' object.

- finalValue (object – 'price')

'FinalValue' is an object describing the total value of awarded contract using the variables in 'Price' object.

- type (enum)

The variable specifies the type of subject of contract. It is an enumerated variable with the following values: *services, goods or construction.*

- size (enum)

The variable specifies the size of the contract. It is an enumerated variable with the following values:

above the threshold, below the threshold, small scale.

- callForTendersPublicationDate (date)

Date of publication of 'call for tenders'. It has NULL value if the call for tenders was not published. The variable is stored in date format.

- bidDeadlineDate (dateTime)

The variable shows the date by which bids have to be submitted. It is not an equivalent to application deadline for 'Restricted' procedure type. The variable is stored in date format.

- documentsDeadlineDate (dateTime)

The variable contains the date until which tender documents or additional information is provided. The variable is stored in date format.

- documentsPayable (bool)

The variable specifies whether there are any fees charged for accessing tender documents or not. It is a Boolean variable having two values: true and false.

- documentsPrice (object – 'price')

'DocumentPrice' is an object describing the price of tender documents using variables in 'Price' object.

- documentsLocation (object – 'address')

This is an object which describes the location where tender documentation can be obtained using variables in 'Address' object.

- contractAwardNoticePublicationDate (date)

The variable shows the publication date of the contract award notice. It also serves as identification if the contract award has been published. It has NULL value if contract award was not published. The variable is stored in date format.

- lastUpdate (date)

Last update of contract data on the source website. The variable is stored in date format.

- awardByGroupBuyers (bool)

The variable specifies if the contract is awarded by a group of bodies. In that case, the leading body is the one described in the 'buyer' object. It is a Boolean having two values: true and false.

- isCentralPurchase (bool)

This variable shows whether the contract was awarded by the central purchasing authority or not. It is a Boolean variable having two values: true and false.

- buyer (object – 'buyer')

'Buyer' is an object identifying the buyer using variables in 'Buyer' object. Buyer could be the contracting authority or another purchasing body.

- onBehalfOf (object – 'buyer')

If the purchase is being made on behalf of another organisation, e.g. a city purchases on behalf of one of its schools, this organisation is identified in this object. 'OnBehalfOf' object uses the 'buyer' object.

- thisPublication (object – ‘publication’)

‘ThisPublication’ is an object describing the current publication using the variables of ‘Publication’.

- otherPublications (object array – ‘publication’)

‘OtherPublication’ is an object array describing a list of relevant publications using the variables in ‘Publication’ object.

- parentContractAwardId (string)

The variable contains the identifier of the superior contract, such as a framework agreement or initiation of dynamic purchasing system (DPS) or construction contract which is being extended. It is in string format.

- parentContractAwardDate (dateTime)

The variable contains the date of the superior contract award. The variable is stored in date format.

- parentContractPublicationUrl (URL)

The variable contains link to the superior contract award. It is stored in URL format.

- administrator (object – ‘body’)

‘Administrator’ is an object specifying the external contract administrator to whom the procedure administration was outsourced - if it was. This object uses the variables in ‘body’ object.

- supervisor (object – ‘body’)

‘Supervisor’ is an object containing information about the external supervisor to whom supervision of the contract fulfilment was outsourced - in if it was.

- specificationsCreator (object – ‘body’)

‘SpecificationsCreator’ is an object containing information about the external body, to which preparation of tender documents was outsourced - if it was.

- lots (object array – ‘lot’)

‘Lots’ is an object array containing information about individual lots using variables in ‘lot’ object. More lots may be connected to a contract.

- candidates (object array – ‘bidder’)

This object array describes the list of candidates who expressed their interest in the contract (typically in a 'Restricted procedure' or 'Negotiated procedure with publication'). It uses the variables in the ‘bidder’ object.

- approachedBidders (object array – ‘bidder’)

This object array describes the list of potential suppliers actively approached or consulted by the contracting authority with information on the contract. It uses the variables in ‘bidder’ object.

- documents (object array – ‘document’)

This object array describes the list of documents relevant to the contract (tender documentation, protocol on bids evaluation etc.) It uses the variables in ‘document’ object.

- cftVersion (int)

The version number of the actual Call for tenders notice, starting incrementally from 1, implicitly assuming that 'otherPublications' object array contains previous/further versions. It is an integer.

- caVersion (int)

The version number of the actual Contract award notice, starting incrementally from 1, implicitly assuming that 'otherPublications' object array contains previous/further versions. It is an integer.

- courtProceedings (URL array)

This array contains the list of links to information on any form of commenced or concluded proceeding related to the contract, if there is any.

- courtInterventions (URL array)

This array contains the list of links to information on any court intervention into the tender/contract.

- npwpReason (enum)

This variable describes the reasons for use of 'negotiated procedure without publication'. It is an enumerated variable with the following values: *no bids, research, technical reasons, artistic reasons, legal reasons, urgency, additional work, work repetition, design contest, commodity market, and previous irregular or unacceptable bids.*

- variantsAccepted (bool)

The variable specifies whether variant offers are accepted or not. It is a Boolean variable.

- deposits (string)

The variable contains the description of deposits required (ideally number and currency, decomposed to price).

- personalRequirements (string)

The variable contains the description of personal requirements on bidders in string format.

- economicRequirements (string)

The variable contains the description of economic requirements on bidders in string format.

- technicalRequirements (string)

The variable contains the description of technical requirements on bidders in string format.

- priorNotification (bool)

The variable specifies whether there were any prior information notices issued or not. It is a Boolean variable having two values: true and false.

- correction (bool)

The variable specifies whether there were any correction or modification notices issued or not. It is a Boolean variable having two values: true and false.

- callForTenders (bool)

The variable specifies whether there was a call for tenders published or not. It is a Boolean variable having two values: true and false.

- contractAward (bool)

The variable specifies whether there was a contract award published or not. It is a Boolean variable having two values: true and false.

- appealBodyName (string)

The variable contains the name of the body to which appeals should be filed. It is stored in string format.

- mediationBodyName (string)

The variable contains the name of the body to which appeals should be filed. It is stored in string format.

- coveredByGPA (bool)

The variable specifies whether the contract is covered by a Government purchasing agreement (GPA) or not. It is a Boolean variable having two values: true and false.

- frameworkAgreement (bool)

The variable specifies whether the contract is awarded as a superior framework agreement or not. It is a Boolean variable having two values: true and false.

- addressOfPerformance (object – ‘address’)

‘AddressOfPerformance’ is an object containing information about the exact address of the contract performance, using the variables in ‘address’ object. These variables describe the postcode, NUTS code, city, street and country of the address of performance.

- isDPS (bool)

This variable shows if the contract either initiates a dynamic purchasing system, or is a purchase via one. It is a Boolean variable having two values: true and false.

- estimatedStartDate (date)

The variable shows the estimated date of contract start. It is stored in date format.

- estimatedCompletionDate (date)

The variable shows the estimated date of contract end. It might be calculated based on stated contract duration if the end date is not available. The variable is stored in date format.

- awardDecisionDate (date)

The variable shows the date of decision on contract award. It is stored in date format.

- ContractSignatureDate (date)

The variable shows the date of contract signature. It is stored in date format.

- mainObject (enum)

The variable contains the main Common Procurement Vocabulary (CPV) code of the subject of contract. It is an enumerated variable. The list of CPV codes is available here: <http://simap.ted.europa.eu/web/simap/cpv>

- additionalObjects (enum array)

The variable contains additional Common Procurement Vocabulary (CPV) codes beside the main CPV code of the subject. It is an enumerated variable. More additional CPV codes may be assigned to a contract. The list of CPV codes is available here: <http://simap.ted.europa.eu/web/simap/cpv>

- funding (object array – ‘funding’)

‘Funding’ is an object array containing information about the funding of the contract using the variables in ‘funding’ object. These variables describe the source, the programme, the amount and proportion of funding.

- awardCriteria (object array – ‘award criterion’)

‘AwardCriteria’ is an object describing the criteria on which the award decision is based. It uses the variables in ‘award criterion’ object. These variables describe the name, weight (on a 0-100 scale) and description of a criterion. It also specifies whether the criterion refers to price or not.

- electronicAuctionUsed (bool)

The variable shows whether the contract is awarded by e-auction or not. It is a Boolean variable having two values: true and false.

‘Lot’ object

This object contains information on individual lots (a contract part) of a contract. Some variables are available on national websites at contract and lot level too. It is preferred to have data at lot level, because this is more precise. For most contracts there is only one lot.

- contractNumber (int)

Identifier number of the contract on the TED website.³ It is in integer format.

- lotNumber (int)

This integer variable specifies the identifier number of a lot within a contract.

- lotTitle (string)

The variable contains the title of the lot in original language in string format.

- lotTitleEnglish (string)

³ See the standard public procurement forms on the TED website (contract award notice, Section V. Award of contract, Contract No.) here: http://simap.ted.europa.eu/documents/10184/49059/sf_003_en.pdf

The variable contains the title of the lot in English in string format.

- lotDescription (string)

The variable contains the description of the lot in original language in string format.

- lotDescriptionEnglish (string)

The variable contains the description of the lot in English in string format.

- lotStatus (enum)

The variable contains information about the status of the lot. It is a categorical variable with the following values: *prepared*, *announced (publicly announced, bids are accepted or negotiated)*, *awarded (awarded and being fulfilled)*, *cancelled*, *finished (fulfilled and paid)*.

- mainObject (enum)

The variable contains the main Common Procurement Vocabulary (CPV) code of the subject of the lot. It is a categorical variable. The list of CPV codes is available here: <http://simap.ted.europa.eu/web/simap/cpv>

- additionalObjects (enum array)

The variable contains additional Common Procurement Vocabulary (CPV) codes of the subject of the lot. It is a categorical variable. More CPV codes may be assigned to a lot. The list of CPV codes is available here: <http://simap.ted.europa.eu/web/simap/cpv>

- estimatedValue (object – 'price')

'EstimatedValue' is an object containing information about the estimated value of a lot using the variables in 'price' object.

- finalValue (object – 'price')

'FinalValue' is an object containing information about the total value of awarded contract lot using the variables in 'price' object.

- addressOfPerformance (object – 'address')

'AddressOfPerformance' is an object containing the exact address of the contract lot performance using the variables in 'address' object.

- estimatedStartDate (date)

The variable shows the estimated start date of the contract. It is stored in date format.

- estimatedCompletionDate (date)

The variable shows the estimated end date of the contract. It might be calculated based on stated contract duration if the end date is not available. The variable is stored in date format.

- awardDecisionDate (date)

The variable shows the decision date for the award of the contract. It is stored in date format.

- ContractSignatureDate (date)

The variable shows the date of the contract signature. It is stored in date format.

- completionDate (date)

The variable shows the real completion date of the contract lot. It is stored in date format.

- cancellationDate(date)

The variable contains the date of the contract lot cancellation. It is stored in date format.

- cancellationReason (string)

The variable contains the stated reason for cancellation of the contract lot, if it was cancelled. It is a string variable

- electronicAuctionUsed (bool)

The variable shows whether the contract is awarded by e-auction or not. It is a Boolean variable having two values: true and false.

- frameworkAgreement (bool)

The variable shows whether the contract is awarded as a superior framework agreement or not. It is a Boolean variable having two values: true and false.

- estimatedNumberOfWinnersInFrameworkAgreement (int)

The variable shows the estimated number of winners of a framework agreement. It is an integer variable.

- isDPS (bool)

This variable shows if the contract either initiates a dynamic purchasing system, or is a purchase via one. It is a Boolean variable having two values: true and false.

- coveredByGPA (bool)

The variable specifies whether the contract is covered by a Government Purchasing Agreement (GPA) or not. It is a Boolean variable having two values: true and false.

- awardCriteria (object array – ‘award criterion’)

‘AwardCriteria’ is an object describing the criteria on which the award decision is based. It uses the variables in ‘award criterion’ object. These variables describe the name, weight (on a 0-100 scale) and description of a criterion. It also specifies whether the criterion refers to price or not. If the decision is based on more criteria, all of them are listed in this object.

- eligibilityCriteria (string)

The variable contains information about the required qualification of a bidder in string format.

- bids (object array – ‘bid’, ‘winning bid’)

‘Bid’ is an object array containing information about received bids using the variables in ‘bid’ and ‘winning bid’ objects. More bids may be listed related to a contract lot. These variables describe the value of the bid, disqualified or not, reason for disqualification, the bidder, bid-related documents.

- funding (object array – ‘funding’)

‘Funding’ is an object array containing information on funding using the variables in ‘funding’ object. These variables describe the source, the programme, the amount and proportion of funding. More funding resources may be connected to one contract.

- bidsCount (int)

This integer variable contains the number of bids received.

- validBidsCount (int)

This integer variable contains the number of bids considered.

- electronicBidsCount (int)

The integer variable contains the number of bids received via electronic means.

‘Bid’ object

This object contains information on the bids received for tenders.

- bidPrice (object – ‘price’)

‘BidPrice’ is an object containing information about the price offered in the bid using variables in ‘price’ object.

- wasDisqualified (bool)

The variable shows whether the bid was disqualified or not. It is a Boolean variable having two values: true and false.

- disqualificationReason (string)

The variable shows the stated reason of the disqualification of bid(s) in string format, if there are disqualified bids.

- bidDocument (object array– ‘document’)

‘BidDocument’ is an object containing information about bid related documents using variables in ‘document’ object.

- bidder (object – ‘bidder’)

‘Bidder’ is an object containing information on bidder using variables in ‘bidder’ object.

‘Winning Bid’ object (Extends ‘Bid’)

This object is an extension of ‘Bid’ object. It is used only if it refers to the winning bid out of all bids. It contains information about the final value, actual payments and evaluation of the completion.

- wasInRequestedQuality (bool)

The variable shows the buyer’s ex post evaluation of the contract completion quality; whether the completion answers the requested quality. It is a Boolean variable having two values: true and false.

- wasFinishedOnTime (bool)

The variable specifies whether the contract was finished on time or not, based on ‘completionDate’ variable or buyer’s ex post evaluation of timely completion. It is a Boolean variable having two values: true and false.

- wasForEstimatedValue (bool)

This variable specifies whether the final payment was in line with the previous estimates or not. It is based on ‘payment’ object and ‘bidPrice’ object or ex post evaluation of contract completion by the buyer. It is a Boolean variable having two values: true and false.

- payments (object array – ‘payment’)

‘Payments’ is an object array containing information on actual payments using variables in ‘Payment’ object. These variables describe the date and amount of payment.

- agreement (object – ‘agreement’)

‘Agreement’ object contains information on the agreement – the specific, actual contract, the document itself - between the issuer and the winner company. It uses the variables in ‘agreement’ object.

- isSubcontracted (bool)

The variable specifies whether the contract was subcontracted or not. It is a Boolean variable having two values: true and false.

- subcontractedProportion (float)

The variable shows the percentage of subcontracted contract volume. The data type of ‘subcontractedProportion’ is float.

- finalValue (object – ‘price’)

‘FinalValue’ is an object describing the price offered by the winner using variables in ‘price’ object.

'Publication' object

This object contains information on the public procurement announcements (call for tenders, contract announcement etc.) referring to the same contract.

- sourceName (enum)

Name of the source of the publication (e.g. TED, DIGIWHIST, Buyer profile). It is an enumerated variable.

- sourceId (string)

Identifier of the contract within the source. Ideally this code should be referred also by assignedContractId but in practice it is possible that different publications are not consistent regarding this. It is a string variable.

- machineReadableURL (bool)

The variable contains the URL of the contract within the source where machine readable data are present. It has NULL value if machine readable data are not available.

- humanReadableURL (URL)

The variable contains the URL of contract on the source where human readable data are present. It has NULL value if human readable data are not available

- publicationDate (date)

The variable contains the date of publication of given announcement version. It is stored in date format.

- dispatchDate (date)

The variable contains the date of dispatch of announcement. It is stored in date format.

- language (enum)

The variable specifies the default language of the data. It is an enumerated variable with ISO 639 language codes available here: http://www.iso.org/iso/home/standards/language_codes.htm

- version (int)

This variable contains the version number of publication on given source. It is an integer.

- valid (bool)

The variable specifies whether the given publication is valid or made obsolete by other listed publications. It is a Boolean variable having two values: true and false.

- included (bool)

The variable specifies whether referred publication data have already been merged into this publication package or not. It is a Boolean variable having two values: true and false.

- releaseTag (enum)

The variable identifies the given type of publication. It is an enumerated variable with values listed here: http://ocds.open-contracting.org/standard/r1_0_RC/en/schema/codelists/#release-tag

'Body' object

This object contains information about a body - as a generic type of legal subject, such as buyer and bidder, subsupplier etc.

- bodyIds (object array – 'body identifier')

The 'bodyIds' object array contains identifier numbers of a body such as VAT number or other unique ID used. The object describes the ID, the type of ID and the scope of ID using the variables in 'Body identifier' object.

- bodyName (string)

This variable contains the name of the body in string format.

- bodyAddress (object – 'address')

This object describes the address of the body using the variables in 'Address' object. These variables contain the raw address in string, the city, postcode, street, NUTS code and country of the address.

- email (string)

This variable contains the e-mail address of the body in string format.

- web (URL)

This variable contains the website of the body in URL format.

- contactPoint (string)

This variable contains the contact point of the body in string format.

- contactName (string)

This variable contains the name of the contact person of the body in string format.

- contactPhone (string)

This variable contains the phone number of the contact of the body in string format.

'Body Identifier' object

The 'body identifier' object uniquely identifies the body within a specific scope.

- id (string)

The variable contains the identifier of the body e.g. a specific VAT ID. It is a string variable.

- idType (enum)

The variable contains the type of the identifier of the body e.g. VAT ID or CZ Business registry etc. The variable is categorical. Its values will be defined during the data collection phase.

- idScope (enum)

The variable specifies the scope under which the identifier is unique. E.g. country code (ISO 3166-1 alpha-2 = two letter), or EU for VAT ID. It has NULL value for global identifiers such as OpenCorporates or Bureau Van Dijk identifiers. The variable is enumerated. Its values will be defined during data collecting phase.

‘Buyer’ object

‘Buyer’ object is an extension of ‘body’ object. It is used if ‘body’ refers to a buyer in the specific case.

- mainActivity (enum)

The variable specifies the main activity of the buyer. It is an enumerated variable with the following values:

general public services, social protection, education, health, environment, public order and safety, housing and community amenities, defence, economic and financial affairs, recreation culture and religion, gas and heat production, gas and oil extraction, coal and other extraction, electricity, water, postal, railway, urban transport, port, airport, other national raw terms.

- buyerType (enum)

The variable specifies the type of the buyer. It is an enumerated variable with the following values:

ministry or any other national or federal authority, including their regional or local sub-divisions; national or federal agency/office; regional or local authority; regional or local agency/office; body governed by public law; European institution/agency or international organization; other.

- isPublic (bool)

The variable specifies whether the buyer is public or not. It is a Boolean variable having two values: true and false.

- isSubsidized (bool)

The variable specifies whether the buyer is subsidized or not. It is a Boolean variable having two values: true and false.

- isSectoral (bool)

The variable specifies whether the buyer is sectoral. It is a Boolean variable having two values: true and false.

‘Bidder’ object

‘Bidder’ object is an extension of ‘body’ object. It is used, if ‘body’ refers to a bidder in the specific case.

- isConsortium (bool)

The variable specifies whether the bidder is an ad hoc consortium of multiple bodies. If it is a consortium, the leading body is specified as the bidder. It is a Boolean variable having two values: true and false.

- subcontractors (object array – ‘body’)

‘Subcontractors’ is an object array specifying all subcontractors, including indirect ones using the variables in ‘Body’ object.

- documents (object array – ‘document’)

‘Documents’ is an object array describing associated documents, such as the final contract or document describing ownership structure. It uses the variables in ‘Document’ object.

‘Funding’ object

This object contains information on the funding sources used in the tender. If none is listed, we implicitly assume that the body uses only its own budget.

- source (string)

The variable contains the name of the funding source in string format.

- euFund (bool)

The variable specifies whether the contract is funded by EU funds or not. It is a Boolean variable having two values: true and false.

- programme (enum)

The variable contains the refined classification of funding sources (such as type of Operational programme etc.) It is an enumerated variable. Its values will be defined during data collecting phase.

- amount (object – ‘Price’)

‘Amount’ is an object containing the amount of contribution from the source using variables in ‘Price’ object.

- proportion (float)

The variable contains the percentage of contribution from the source in a float data format.

‘Award Criterion’ object

This object contains information on the award criteria used in the tender. If the lowest price is used, only one criterion is recorded.

- name (string)

The variable contains the name of the award criterion used. It is always 'price' if the award criteria are defined as 'lowest price' in the call for tenders. It is a string variable.

- weight (float)

The variable contains the weight of a given criterion on scale 0-100. The data type of this variable is float.

- description (string)

The variable contains the broader description of the criterion in string format.

- priceRelated (bool)

The variable specifies whether the criterion is directly related to price (monetary measurable) or not. It is defined based on the occurrence of keywords such as e.g. price, lifetime costs, interest etc. It is a Boolean variable having two values: true and false.

‘Qualification’ object

This object describes structured information on qualification criteria and requirements, including deposits.

- type (enum)

The variable contains the type of requirement. It is an enumerated variable with the following values: *technical personal, economic*.

- description (string)

The variable describes the requirement in string format.

‘Payment’ object

This object contains information on payment, the actual transfer of money related to the tender.

- paymentDate (date)

The variable contains the date of payment. It is stored in date format.

- amount (object – ‘price’)

‘Amount’ is an object showing the amount paid without VAT using the variables in ‘Price’ object.

‘Document’ object

This object contains metadata on tender related documents.

- title (string)

The variable shows the title of the document in string format.

- type (enum)

The variable contains the type of the document e.g. agreement, tender etc. It is an enumerated variable with the following values:

project plan, market research, specifications, qualification documents, call for tenders, additional information, candidates number limitation protocol, qualifications evaluation protocol, bidder negotiation protocol, ownership structure, objection, tender, evaluation committee protocol, tenders evaluation protocol, best tender selection notice, contractor agreement, agreement amendment, buyer report, buyers grouping agreement, agreement appendix, national term .

- URL (URL)

The variable contains the URL of the document in URL format.

- publicationDateTime (date)

This variable shows the publication date and time of the document in dateTime format.

- signatureDate (date)

This variable shows the signature date of the document in date format.

- otherVersions (object array – ‘document’)

‘Other versions’ is an object array describing other versions of the document using variables in ‘Document’ object. The actual document is recognized by highest publicationDateTime.

‘Agreement’ object

This object extends ‘document’ object with information on annexes, amendments.

- appendices (object array – ‘Document’)

‘Appendices’ is an object array describing the list of annexes using ‘Document’ object.

- amendments (object array – ‘Document’)

‘Amendments’ is an object array describing the list of amendments using ‘Document’ object.

'Price' object

This object contains detailed information on prices.

- netAmount (double)

This variable contains the price without VAT in double format.

- vat (float)

This variable contains the VAT percentage. It has NULL value if no VAT is paid in float format.

- currency (enum)

The variable contains the currency of the price. It is an enumerated variable with values based on ISO 4217 standard available here: http://www.iso.org/iso/home/standards/currency_codes.htm

- netAmountEur (double)

The variable contains the price in EUR without VAT in double format.

- unitPrice (double)

The variable contains the price of a single unit in the same currency, if the price is given for a single unit. The data format is double.

- unitNumber (int)

This integer variable shows the number of units if the price is given for a single unit.

- unitType (string)

The variable describes the unit if the price is given for a single unit (e.g. EUR/KG). It is a string variable.

- minPrice (double)

The variable shows the minimum reachable value of 'netAmount' variable in double format if not an exact price but an interval is given.

- maxPrice (double)

The variable shows the maximum reachable value of 'netAmount' variable in double format if not an exact price but an interval is given.

'Address' object

'Address' object contains detailed information about locations.

- rawAddress (string)

This variable contains the whole address in a raw, unstructured string format.

- street (string)

The variable contains the street and numbers of the address in a string format

- city (string)

The variable contains the city or town of the address in a string format

- postcode (string)

The variable contains the postcode of the address in a string format.

- nuts (enum)

The variable contains the NUTS code of the region. It is an enumerated variable containing official NUTS codes as values. The list of values is available here:

[http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NU
TS_2013L&StrLanguageCode=EN&StrLayoutCode=HIERARCHIC](http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NU
TS_2013L&StrLanguageCode=EN&StrLayoutCode=HIERARCHIC)

- country (enum)

The variable shows the country in the address. It is an enumerated variable containing values based on 'ISO 3166-1 alpha-2 = two letter' standard which is available here:

http://www.iso.org/iso/country_codes

2. Company data

DIGIWHIST's database will also include detailed company data, which is necessary to be able to assess company level risks of corruption and favouritism thoroughly. Unlike public procurement, public authority data or public official data, company data will be purchased from a private provider. Private providers automatically as well as manually create their databases from fragmented national company registries and warehouses, almost exclusively relying on official primary sources. Consequently, data availability will depend on the data scope of private providers, however, there are approximately 26 million companies in the 35 jurisdictions included (based on Eurostat). Although republishing of contract related company information will not be possible, derived indicators can be displayed. Company data consists of four main categories, 1) Registry data, 2) Financial data, 3) Ownership data and 4) Managers/directors data. Each of these is discussed in more detail below. Company data will be stored in an object based database structure, however the exact structure will depend on the data provider's input.

Registry data

- Company name

Definition: Name of the company.

Type: string

- Company also known name (alternative name(s))

Definition: Alternative name of the company.

Type: array - string

- HQ address: Street, no., building etc

Definition: Address of the company's headquarters.

Type: string

- HQ postcode

Definition: The postal code of the headquarters.

Type: String

- HQ city

Definition: The city/town of the headquarters.

Type: Sting

- HQ county (NUTS3)

Definition: The NUTS3 region code of the headquarters.

Type: String

- HQ country ISO code

Definition: Country ISO code of the company. (ISO 3166-1 alpha-2 = two letter)

Type: enum

- Unique ID number (internal to the database)

Definition: The unique identifier of the organization.

Type: string

- EuropeanVATnumber

Definition: The European VAT number of the organization.

Type: string

- National company ID number

Definition: National identifier of the organization. (e.g. CZ business registry)

Type: string

- National tax ID (if different from national ID number)

Definition: National tax ID of the company.

Type: string

- National legal form

Definition: The legal form of the organization.

Type: string

- Date of incorporation

Definition: Date of incorporation of the organization.

Type: date

- Date of dissolution/bankruptcy/etc

Definition: Date of dissolution/bankruptcy etc. of the organization.

Type: date

- Dissolution type (codelist of types)

Definition: Type of dissolution of the company. (codelist of types)

Type: enum

- Reporting basis (e.g. consolidated accounts)

Definition: Reporting bases of the company. (e.g. consolidated accounts)

Type: string

- Main sector (nace rev2)

Definition: Statistical classification of the organization based on economic activities. (NACE Rev.2)

Type: enum

- Publicly quoted company (yes/no)

Definition: Specifies whether the organization is publicly quoted or not.

Type: bool

- Company size category (e.g. small, medium, large)-include precise definitions too

Definition: Classification of company based on size. (e.g. small, medium, large)

Type: enum

- Number of employees

Definition: Number of employees of the company.

Type: integer

- Main foreign country

Definition: Main foreign country of the company. (ISO 3166-1 alpha-2 = two letter)

Type: enum

- Main production sites' location

Definition: Main production sites' of the company. (NUTS3)

Type: enum

- Main distribution sites' location

Definition: Main distribution sites' of the company. (NUTS3)

Type: enum

Financial data

- Total Revenues/Turnover

Definition: Total revenues/turnover of the company.

Type: float

- Operating revenue

Definition: The company's income derived from sources related to the company's everyday business operations.

Type: float

- Financial revenue

Definition: The company's income derived from interest, rent and other incomes earned in owning or renting an asset or property.

Type: float

- Other Revenues

Definition: Other revenues of the company. This is revenue derived from any source other than the company's operations.

Type: float

- Material costs

Definition: Material costs of the company. This is the cost of materials used for production.

Type: float

- Personnel Expenses/cost of employees

Definition: Salary and wage (including fringe benefits) costs of the company.

Type: float

- Total Depreciation, Amort. & Depl.

Definition: Total depreciation, amortization and depletion of the company.

Type: float

- Research & Development Expenses

Definition: All expenses of the company related to research and development.

Type: float

- Financial Expenses(including interests paid)

Definition: Value of expenditures incurred by the organization during the reporting period due to interest, fees, and commissions incurred on the organization's liabilities, including any client deposit accounts held by the organization, borrowings, subordinated debt, and other financial liabilities during the reporting period.

Type: float

- Financial P/L

Definition: The company's net profits and loss.

Type: float

- Profit before Tax

Definition: The company's profits before paying corporate income tax, but deducting all expenses (operating and financial).

Type: float

- Taxes paid on corporate income

Definition: Corporate income tax of the company.

Type: float

- Profit after Tax

Definition: Net amount earned after deduction of all tax related expenses.

Type: float

- Retained Profit Brought Forward

Definition: Retained profit brought forward by the company.

Type: float

- Retained Profit Carried Forward

Definition: Retained profit carried forward by the company.

Type: float

- Dividends paid

Definition: Dividends paid by the company.

Type: float

- EBITDA

Definition: The company's earnings before interest, taxes, depreciation and amortization.

Type: float

- ROA

Definition: Return on assets equals the company's annual earnings divided by its total assets.

Type: float

- Capital Expenditures

Definition: Capital expenditures (CAPEX) are funds used by a company to acquire or upgrade fixed, physical, non-consumable assets, such as buildings and equipment or a new business.

Type: float

- Net total investment

Definition: The amount spent by a company or an economy on capital assets, or gross investment, less depreciation.

Type: float

- Net Cash

Definition: The company's total cash minus total liabilities.

Type: float

- Total tangible assets

Definition: Company assets that have a physical form; includes both fixed assets, such as machinery, buildings and land, and current assets, such as inventory.

Type: float

- Total intangible assets

Definition: Includes company assets that are not physical in nature, including e.g. corporate intellectual property (items such as patents, trademarks, copyrights, business methodologies), goodwill and brand recognition.

Type: float

- Total assets

Definition: The total amount of all gross investments, cash and equivalents, receivables, and other assets as they are presented on the company's balance sheet.

Type: float

- Total liabilities

Definition: The aggregate of all debts an individual or company is liable for.

Type: float

- Total short term debt

Definition: Any debt incurred by a company that is due within one year.

Type: float

- Total long term debt

Definition: Any debt incurred by a company that is due in over one year.

Type: float

Ownership data

- No of recorded shareholders

Definition: Number of recorded shareholders of a company.

Type: integer

- Shareholder - Name

Definition: Name of the company's shareholder(s).

Type: string

- Shareholder – Salutation

Definition: Salutation of the company's shareholder(s).

Type: enum

- Shareholder - First name

Definition: First name of the company's shareholder(s).

Type: string

- Shareholder - Last name

Definition: Last name of the company's shareholder(s).

Type: string

- Shareholder - ID number

Definition: ID of the company's shareholder(s).

Type: string

- Shareholder - type (individual, legal entity, bank, etc)

Definition: Type of the company's shareholder(s).

Type: enum

- Shareholder - Legal Entity Identifier

Definition: Legal entity of the company's shareholder(s).

Type: enum

- Shareholder - Country ISO code

Definition: ISO code of the company's shareholder country. (ISO 3166-1 alpha-2 = two letter)

Type: enum

- Shareholder - State or province (in US or Canada)

Definition: State or province of the company's shareholder(s).

Type: string

- Shareholder – City

Definition: City/town of the company's shareholder(s).

Type: string

- Shareholder - Potentially in WorldCompliance database

Definition: Specifies whether the company's shareholder(s) is included in WorldCompliance database.

Type: bool

- Shareholder - NACE Rev. 2, Core code

Definition: Sector of the company's shareholder(s). (NACE Rev.2)

Type: enum

- Shareholder - Operating revenue (Turnover)

Definition: Operating revenue of the company's shareholder(s).

Type: float

- Shareholder - Total assets

Definition: Total assets of the company's shareholder(s).

Type: float

- Shareholder - Number of employees

Definition: Number of employees of the company's shareholder(s).

Type: integer

- Shareholder - Company taking liability

Definition: Shareholder company taking liability for the company's debt.

Type: bool

- Shareholder - Direct %

Definition: Direct share of the company's shareholder(s).

Type: float

- Shareholder - Total %

Definition: Total share of the company's shareholder(s).

Type: float

- Shareholder - Information source

Definition: Source of information regarding the company's shareholder.

Type: string

- Shareholder - Information date

Definition: Last update of shareholder(s) information.

Type: date

- Shareholder - Also a manager

Definition: Specifies whether the shareholder is also a manager of the company or not.

Type: bool

- Shareholder - starting date

Definition: Shareholder's starting date.

Type: date

- Shareholder - Closing date

Definition: Shareholder's closing date.

Type: date

Managers/directors data

- Number of Directors

Definition: Number of directors of the company.

Type: integer

- Number of current Directors

Definition: Number of current directors of the company.

Type: integer

- Number of Previous Directors

Definition: Number of previous directors of the company.

Type: integer

- Full name

Definition: Full name of the company's directors

Type: array - string

- Unique Contact Identifier no.

Definition: Unique contact identifier of the company's managers/directors.

Type: array - string

- In Worldcompliance database

Definition: Specifies whether manager(s)/director(s) are included in the Worldcompliance database.

Type: bool

- Position/Job Title

Definition: Position/job title of the director/manager.

Type: string

- Appointment Date

Definition: Appointment date of the manager(s)/director(s).

Type: date

- Resignation Date

Definition: Date of resignation of the manager(s)/director(s).

Type: date

- Gender

Definition: Gender of the manager(s)/director(s).

Type: bool

- Date of Birth

Definition: Date of birth of the manager(s)/director(s).

Type: date

- Nationality

Definition: Nationality of the manager(s)/director(s).

Type: enum

- Country

Definition: Country of the manager(s)/director(s).

Type: enum

- Also a shareholder? (yes/no)

Definition: Specifies whether the manager(s)/director(s) is/are shareholder(s) or not.

Type: bool

3. Public authority data

DIGIWHIST will compile a comprehensive registry of public authorities (contracting authorities) for each jurisdiction. When feasible, these organizational units generally include central administration bodies (such as ministries, departments and depending units); autonomous agencies; regions, provinces or municipalities; state-owned enterprises and publicly-funded organizations. Similarly to the public procurement database, public authority data will also be an object based dataset; however, the final format will only be concluded after a thorough mapping of (feasibly) available information.

- ID

Definition: Internally generated ID.

Type: Integer

- Name

Definition: Name of the public authority.

Type: String

- Jurisdictional level

Definition: Jurisdictional level of the public authority. (e.g. central administration, public body etc.)

Type: Enum

- Parent contract authority ID

Definition: ID of the parent subject of the given authority (if any).

Type: String

- Country specific code

Definition: Country specific ID of the public authority.

Type: String

- Activity type

Definition: Activity type of the public authority. (e.g. Health etc.)

Type: String

- Type

Definition: Type of the public authority. (e.g. Regional etc.)

Type: String

- Country

Definition: Country code of the public authority. ([ISO 3166-1: http://www.iso.org/iso/country_codes](http://www.iso.org/iso/country_codes))

Type: Enum

- Address

Definition: Address of the public authority.

Type: String

- Town

Definition: City/town of the public authority.

Type: String

- Postcode

Definition: Postcode of the public authority.

Type: String

- Contact person

Definition: Contact person of the public authority.

Type: String

- Phone

Definition: Phone number of the public authority.

Type: String

- Fax

Definition: Fax number of the public authority.

Type: String

- E-mail

Definition: E-mail address of the public authority.

Type: String

- Web

Definition: Web address of the public authority.

Type: String

Budget data

In addition to general public authority data, DIGIWHIST also collects financial information on public authorities, such as overall annual spending, but also personnel costs, revenues and indebtedness if such information is obtainable. It will be stored in an object based database structure; however the exact database scheme can be only finalized when the scope of available data is properly mapped.

- ID

Definition: Internal ID used for the given budget item.

Type: Integer

- Contracting authority's ID

Definition: Internal ID used for the public authority.

Type: Integer

- Amount

Definition: Amount of the given budget item.

Type: Float

- Type

Definition: Type of the given budget item.

Type: Enum

- Currency

Definition: Currency of the given budget item.

Type: Enum

- Amount in EUR

Definition: Amount in EUR of the given budget item.

Type: Float

- Year

Definition: Year of the given budget item.

Type: Float

- Item

Definition: Classification of the given budget item. (e.g. the standard used in IFRS)

Type: Enum

- Item Classification

Definition: Classification of the budget item. (e.g. IFRS)

Type: Enum

4. Public officials

DIGIWHIST also collects data on public officials in charge of each contracting authority. Besides the managers of public authorities, data on members of national, regional and municipal level legislative bodies will be also included when feasible. Data on public official will be stored in an object based database scheme as well. However, it is important to note that the scope of online published data is much narrower than the data stored in the database due to protection of personal data. Derived variables characterising higher order units such as countries will be released, but never data relating to specific individuals in line with data protection and privacy concerns discussed in the Description of Work.

- ID

Definition: Internal ID used for the public official.

Type: Integer

- Contracting authority's ID

Definition: Contracting authority ID of the public official.

Type: Integer

- Name

Definition: Name of the public official.

Type: String

- Position

Definition: Position of the public official.

Type: Enum

- Started

Definition: Start date of the public official's position.

Type: Date

- Ended

Definition: End date of the public official's position.

Type: Date

- Political membership

Definition: Political party membership of the public official while holding position.

Type: Enum

III. PART. AVAILABILITY OF PUBLIC PROCUREMENT DATA

This section shows the availability of public procurement data in terms of exact variables that were defined in previous sections. It is important to note that Table 1. shows a theoretic maximum of available data; it shows countries where the given information should be available according to national regulation but we will not have reliable information about the completeness of these data until the beginning of the data retrieving phase.

Table 1: Data availability in published tender documents across countries

Variable names	Austria	Belgium	Bulgaria	Cyprus	Croatia	Czech Rep.	Germany	Denmark	Estonia	Spain	Finland	France	Greece	Hungary	Ireland	Italy	Lithuania	Luxembourg	Latvia	Malta*	Netherlands	Poland	Portugal	Romania	Sweden	Slovenia	Slovakia	United Kingdom	Norway	Switzerland	Iceland	Serbia	Georgia	Armenia	EC		
	title	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
procedureType	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
npwpReason	X		N/A			X	X	X						X				X				X	N/A		N/A	X									X		
description	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
type	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
size	X		X	X	X	X	X	X	X	X		X		X					X			X	X	X	N/A		X	X		X						X	
addressOfPerformance	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
variantsAccepted	X	X	X		X				X	X	X	X	X	X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
deposits	X	X	X	X			X			X		X	X	X			X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
electronicAuctionUsed	X	X	X	X	X	X	X		X	X		X	X	X	X		X	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
frameworkAgreement	X	X	X	X		X	X		X	X		X	X	X			X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
awardCriteria	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
mainObject	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
assignedAnnouncementIds	X	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
assignedContractIds	X	X	X	X		X			X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
IsDPS	X	X	X			X	X					X		X			X	X			X	X	X	N/A		X	X					X				X	
callForTendersPublicationDate	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
bidDeadlineDate	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
documentsDeadlineDate	X	X	X	X		X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
contractAwardNoticePublicationDate	X	X	X	X		X			X	X		X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
lastUpdate								X																										X	X		
estimatedStartDate	X	X		X		X	X		X		X		X	X			X		X		X		X	X		X	X	X	X	X	X	X	X	X	X	X	X
estimatedCompletionDate	X	X	X	X		X	X	X	X	X			X	X			X	X	X		X		X	X		X	X	X	X	X	X	X	X	X	X	X	X
awardDecisionDate	X	X	X	X		X			X	X		X	X	X	X	X	X		X	X	X	X	X	X		X	X		X				X			X	
estimatedCompletionDate			X											X																							
cancellationDate	X	X	N/A	X		X		X		X			X	X			X		X		X						X	X					X			X	

Table 3: Data availability in published tender documents across countries (part III)

	Variable names	Austria	Belgium	Bulgaria	Cyprus	Croatia	Czech Rep.	Germany	Denmark	Estonia	Spain	Finland	France	Greece	Hungary	Ireland	Italy	Lithuania	Luxembourg	Latvia	Malta*	Netherlands	Poland	Portugal	Romania	Sweden	Slovenia	Slovakia	United Kingdom	Norway	Switzerland	Iceland	Serbia	Georgia	Armenia	EC		
Bidder/Winner information	bodyName	X	X	X	X		X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	bodyAddress	X	X	X	X		X	X		X	X		X	X	X				X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X
	bodyIds			X							X					X	X	X			X				X									X	X			
	bodyName (other bidders)			X											X						X											X	X					
	bodyAddress (other bidders)														X																				X			
	bodyIds (other bidders)			X																															X			
	bidderLimitation	X	X				X	X		X				X	X				X	X	X		X	X		X		NA										X
	estimatedNumberOfWinnersInFrameworkAgreement			X										X		X			X		X					X												X
	isConsortium																					X						NA										NA
	isSubcontracted		X	X			X			X			X		X				X	X	X		X		X	X	X	X	X	X				X				X
	bodyName (subcontractor)			NA																	X				NA													
subcontractedProportion		X	NA			X			X			X		X				X		X						X	X											X
documents																					X																X	
Bids	bidsCount	X	X	X	X		X			X	X		X		X			X	X	X	X	X	X	X	X	X	X		NA	X		X	X	X	X	X	X	
	validBidsCount												X															X										
	electronicBidsCount		X				X								X				X				X				X									X	X	
	wasDisqualified																						X															
	disqualificationReason																																					
Prices	estimatedValue	X	X	X	X		X	X	X	X	X		X	X	X			X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	finalValue		X	X	X		X		X	X	X		X	X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	bidPrice														X						X															X	X	
	payments (at contract completion)			X											X									X												NA		
	maxBid		X				X				X				X				X	X	X		X							X		X	X	X	X	X	X	X
minBid		X				X				X				X				X	X	X		X							X		X	X	X	X	X	X	X	

Table 4: *Data availability in published tender documents across countries (part IV)*

	Variable names	Austria	Belgium	Bulgaria	Cyprus	Croatia	Czech Rep.	Germany	Denmark	Estonia	Spain	Finland	France	Greece	Hungary	Ireland	Italy	Lithuania	Luxembourg	Latvia	Malta*	Netherlands	Poland	Portugal	Romania	Sweden	Slovenia	Slovakia	United Kingdom	Norway	Switzerland	Iceland	Serbia	Georgia	Armenia	EC			
Cancellation/correction	lotStatus (value: cancelled)	X	X	X			X			X	X	X	X	X	X		X	X		X	X	X	N/A	X			X	X		X	X	X	X			X			
	cancellationReason	X		N/A	X		X		X		X			X	X			X		X		X	N/A	X			X	X		X	X		X						
	correction	X	X	X	X					X	X	X	X	X	X		X	X		X	X	X	X	X	X	X	X		X	X		X	N/A			X			
Other information	administrator														X																						X		
	supervisor																																						
	specificationsCreator														X													N/A											
	courtProceedings			X										X	X										X														
	courtInterventions			X											X					X				X															
	appealBodyName	X	X	X	X		X	X	X	X	X	X		X	X	X			X	X	X	X	X	X	X	X				X		X				X	X		
mediationBodyName		X											X	X				X	X	X	X														X	X			

X= information is available;

N/A=information availability could not be confirmed

Blank=information is not available

IV. PART. DIGIWHIST AND OPEN CONTRACTING DATA STANDARD COMPARISON

This section provides a documentation of the similarities and differences between the data collected in DIGIWHIST and Open Contracting Data Standard's data structure. Clear correspondence between DIGIWHIST public procurement data template and the quickly spreading global Open Contracting Data Standard (OCDS) is highly desirable to strengthen the global movement and to show ways of further development by good example. Table xx shows that on one hand DIGIWHIST covers a wider range of information regarding European public procurement than OCDS due to more extensive data provision requirements in most jurisdictions in Europe. On the other hand, there are only few variables that are available in OCDS but not in DIGIWHIST. All in all, there is a one to one translation between fields in the DIGIWHIST data template and OCDS with a few exceptions only. Such correspondence allows for taking into account European specificities, while also directly tapping into an incipient global movement.

Table 5: Comparison of DIGIWHIST and Open Contracting Data Standard variables

DIGIWHIST attribute name	OCDS equivalent	Comment
CONTRACT		
id	id	
announcementType	release.tag	
country		
assignedContractIds	-	
assignedAnnouncementIds	<u>document.id</u>	
relatedAnnouncementIds	-	
title	tender.title	
titleEnglish	tender.title	With language mutation
procedureType	tender.procurementMethod	DIGIWHIST's list is more comprehensive
nationalProcedureType	-	
description	tender.description	
descriptionEnglish	tender.description	With language mutation
bidderLimitation	-	
estimatedValue	tender.value	
finalValue	contract.value	
type	-	
size	-	
callForTendersPublicationDate	within documents	
bidDeadlineDate	tender.tenderPeriod	

documentsDeadlineDate	-	
documentsPayable	-	
documentsPrice	-	
documentsLocation	-	
contractAwardNoticePublicationDate	within documents	
lastUpdate	within documents	
awardByGroupBuyers	-	
isCentralPurchase	-	
buyer	tender.buyer	
onBehalfOf	-	
thisPublication	within documents	
otherPublications	within documents	
parentContractAwardId	-	
parentContractAwardDate	-	
parentContractPublicationUrl	-	
administrator	-	
supervisor	-	
specificationsCreator	-	
lots	awards	Slightly different approach
candidates	-	
approachedBidders	-	
documents		Yes, on multiple places (separate documents for contract, planning etc.)
cftVersion		OCDS has more elaborate approach for publishing versions
caVersion		OCDS has more elaborate approach for publishing versions
courtProceedings	-	
courtInterventions	-	
npwpReason	procurementMethodRationale	OCDS has a more general version
variantsAccepted	-	
deposits	-	
personalRequirements	tender.eligibilityCriteria	DIGIWHIST is more detailed
economicRequirements	tender.eligibilityCriteria	DIGIWHIST is more detailed
technicalRequirements	tender.eligibilityCriteria	DIGIWHIST is more detailed
priorNotification	within documents	
correction		OCDS has more elaborate approach for publishing

		versions
callForTenders	release.tag	
contractAward	release.tag	
appealBodyName	-	
mediationBodyName	-	
ANY NON OBJECT LOT LEVEL VARIABLE - see those below		
coveredByGPA	-	
frameworkAgreement	-	
addressOfPerformance	-	
isDPS	-	
estimatedStartDate	award.contractPeriod	
estimatedCompletionDate	award.contractPeriod	
awardDecisionDate	-	
ContractSignatureDate	Contract.dateSigned	
mainObject	can be partially transformed into tender.items, award.items	
additionalObjects	can be partially transformed into tender.items, award.items	
funding	similar to planning.budget	
awardCriteria	tender.awardCriteria	DIGIWHIST is more general allowing for various criteria with weights.
electronicAuctionUsed	tender.submissionMethod	OCDS is more general.
LOT		
contractNumber	-	
lotNumber	-	
lotTitle	award.title	
lotTitleEnglish	award.title (en)	
lotDescription	award.description	
lotDescriptionEnglish	award.description (en)	
lotStatus	tender.status	DIGIWHIST is lot based, since more detailed. It uses different codelist keywords for the same content
mainObject	tender.items	Partially
additionalObjects	tender.items	Partially
estimatedValue	tender.value	
finalValue	contract.value	
addressOfPerformance	-	
estimatedStartDate	-	

estimatedCompletionDate	-	
awardDecisionDate	award.date	
ContractSignatureDate	contract.dateSigned	
completionDate	-	
cancellationDate	-	
cancellationReason	-	
electronicAuctionUsed	tender.submissionMethod	OCDS is more general.
frameworkAgreement	-	
estimatedNumberOfWinnersInFrameworkAgreement	-	
isDPS	-	
coveredByGPA	-	
awardCriteria	tender.awardCriteria	DIGIWHIST is more general allowing for various criteria with weights.
eligibilityCriteria	is in documents	
bids	contract.tenderers	
funding	is in budget	Similar, not the same
bidsCount	tender.numberOfTenderers	DIGIWHIST is more detailed
validBidsCount	tender.numberOfTenderers	DIGIWHIST is more detailed
electronicBidsCount	tender.numberOfTenderers	DIGIWHIST is more detailed
CONTRACT PUBLICATION		
sourceName	-	
sourceId	-	
machineReadableURL	-	
humanReadableURL	-	
publicationDate	extended version of release.date	
dispatchDate	extended version of release.date	
language	document.language	
version	-	
valid	-	
included	-	
releaseTag	release.tag	
FUNDING		
source	similar to budget.source	
euFund	-	
programme	-	
amount	similar to budget.amount	

proportion	-	
AWARD CRITERION		This entity can be translated into OCDS's tender.awardCriteria with some information loss.
name	-	
weight	-	
description	-	
priceRelated	-	
QUALIFICATION		
type	-	
description	-	
BODY		
bodyIds	organization.identifier, additionalIdentifiers	
bodyName	<u>organization.name</u>	
bodyAddress	organization.Address	
email	organization.contactPoint.email	
web	organization.contactPoint.url	
contactPoint	-	
contactName	<u>organization.contactPoint.name</u>	
contactPhone	organization.contactPoint.telephone	
BODY IDENTIFIER		
id	<u>organization.identifier.id</u>	
idType	organization.scheme	
idScope	-	
BUYER		
mainActivity	-	
buyerType	-	
isPublic	-	
isSubsidized	-	
isSectoral	-	
BIDDER		
isConsortium	-	

subcontractors	-	
documents	similar to award.documents	
BID		
bidPrice	award.value	For winning bid only
wasDisqualified	-	
disqualificationReason	-	
bidDocument	-	
bidder	contract.tenderers, award.suppliers	
WINNING BID		
wasInRequestedQuality	-	
wasFinishedOnTime	-	
wasForEstimatedValue	-	
payments	similar to Implementation	
agreement	contract.documents	
isSubcontracted	-	
subcontractedProportion	-	
finalValue	contract.value	
PAYMENT		
paymentDate	contract.implementation.transaction.date	
amount	contract.implementation.transaction.amount	
DOCUMENT		
title	document.title	
type	document.documentType	
URL	document.url	
publicationDateTime	document.datePublished	
signatureDate	-	OCDS has this for contracts only.
otherVersions	can be translated to document.dateModified	
AGREEMENT		
appendices	-	
amendments	-	

PRICE	tender.items	
netAmount	award.value.amount	
vat	-	
currency	award.value.currency	
netAmountEur	-	
unitPrice	corresponds to tender.items	
unitNumber	corresponds to tender.items	
unitType	corresponds to tender.items	
minPrice	-	
maxPrice	-	
	http://ocds.open-contracting.org/standard/r/1_0_RC/en/schema/reference/#address	
ADDRESS		
rawAddress	-	
street	address.streetAddress	
city	address.locality	
postcode	address.postalCode	
nuts	can be transformed to address.region	
country	can be transformed to address.countryName	

Table 6: Variables that are available in OCDS's standard but not in DIGIWHIST

OCDS variable	comment
release.ocid	Can be generated based on contract.id
release.date	DIGIWHIST uses contract.publications structure instead
release.initiationType	Can be derived from contract.procedureType
release.planning.rationale	DIGIWHIST will have some of these in contract.document
release.planning.documents	DIGIWHIST only has documents attached directly to contract.
release.planning.budget	DIGIWHIST has contract.funding entity which is slightly less detailed.
tender.items	This is implemented within contract.lot.bid.price, lot.mainObject,lot.additionalObjects.
tender.minValue	DIGIWHIST only has one estimated value, as follows from procurement directive.
tender.awardCriteriaDetails	DIGIWHIST has more detailed approach for mapping individual criteria
tender.submissionMethod	Can partially be replaced by contract.electronicAuctionUsed
tender.submissionMethodDetails	
tender.enquiryPeriod	
tender.hasEnquiries	Can be filled using documents (type "Additional information")
tender.awardPeriod	
numberOfTenderers	
organization.contactPoint.faxNumber	
organization.identifier.legalName	Similar to organization.Name
document.description	
document.format	
milestones	Can be partially filled based on documents and publications.
ammendment	
award.title	
contract.implementation.transaction.source	