Deliverable 6.1
DRAFT*

Public Procurement, Companies, and Contracting Networks

Sources, Structure and Documentation

Zdravko Veljanov and Mihaly Fazekas

DATA REPORT

Funded by the European Union

NET-ROL has received funding from the European Union's HORIZON-CL2-2024-DEMOCRACY-01 under grant agreement No 101177405

* awaiting approval by the Commission

NET-ROL
Networks and the Rule of Law: Uncovering Socio-Economic Outcome

NET-ROL has received funding from the European Union's HORIZON-CL2-2024-DEMOCRACY-01 under grant agreement No 101177405

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Suggested citation
Veljanov, Zdravko and Mihaly Fazekas. (2026). *Public Procurement, Companies, and Contracting Networks Sources, Structure and Documentation.* NET-ROL Data Report No.3.

Reviewer: Fabio Padovano

Date: 28.08.2026
Authors: Zdravko Veljanov and Mihaly Fazekas
NET-ROL deliverable no: 6.1
Contact person: FazekasM@ceu.edu

Executive Summary

The report documents Deliverable 6.1 of the NeT-RoL project: an integrated dataset covering public procurement, sanctions from judicial decisions, blacklists, and network data connecting buyers and suppliers. The deliverable is produced under Work Package (WP) 6 and fulfills the data collection subtasks defined in the proposal – public procurement contracts database (subtask 6.2.1) and the collection of court and debarment decisions related to public procurement (subtask 6.2.2).

The procurement component compiles contract level data for eight priority countries: Albania, Austria, Bulgaria, Czech Republic, Germany, Hungary, Italy, and Poland. It builds on and extends the administrative infrastructure created by the Horizon 2020 DIGIWHIST project, CEU's Global Corruption Observatory project, and the ongoing RESPOND project. The data includes improved historical coverage and data quality for recent years in the project's target countries. Each national dataset contains a harmonized and standardized structure, allowing for cross-country comparisons, and is accompanied by a data dictionary (see Annex) document the included variables and their definition.

Additionally, the public procurement datasets integrate a sanction component from judicial decisions or blacklisted component from debarment lists. Sanctions and/or blacklisted data have been collected and merged into the public procurement data for the Czech Republic, Albania, Austria and Hungary. Data for Bulgaria is also collected. Collections for the remaining countries is ongoing.

The network component represents the public procurement data as a supplier-buyer contracting network. For each country an edgelist is constructed in which each edge links contracting authority (buyer) to an economic operator (supplier) weighted by the number of contracts and their total value. These structures support the social network analysis in the proposal.

The datasets are released and published in an open, documented format with accompanying data dictionaries on Mendeley [1] ensuring reproducibility and reuse.

[1] https://data.mendeley.com/datasets/nfhtfwn5nk/1

Introduction and Scope

The purpose of Work Package 6 is to assemble and extend large-scale administrative data on public spending and to link them to companies whose interactions can be studied as networks. The work draws on, builds and improves in terms of (historical) scope and quality on three earlier projects: HORIZON 2020 DIGIWHIST, CEU's Global Corruption Observatory and the RESPOND project.

Deliverable 6.1 is the first data release of the work package. It contains three integrating parts that reflect the tasks of the project:

1. A public procurement contracts database for eight priority countries (Albania, Austria, Bulgaria, Czech Republic, Germany, Hungary, Italy, and Poland) harmonized to a standard schema and extended in historical coverage and data quality to existing datasets;
2. A collection of court and debarment decisions related to public procurement, such as national arbitration court rulings or national debarment lists[2] that are merged into to the public procurement data at the level of economic operators – bidders; and
3. A network representation of the public procurement data as a bipartite supplier-buyer contracting networks.[3]

This report describes the sources, structure, and content of each component. It reports the descriptive statistics and provides links to the published datasets. Detailed variable-level documentation is provided in the Annex.

[2] A debarment list is a publicly maintained register of companies and/or individuals formally barred, either temporarily or permanently, from bidding on public contracts, typically due to bid-rigging, corruption or other serious professional misconduct.

[3] A network representation of public procurement data in which two distinct type of nodes, contract authorities (or buyers) and economic operators (or bidders/suppliers) are linked by edges that represent the awarding contract and contact between the entities. Because the ties only connect buyers with bidders, the network representation is bipartite. The structure allows analysis of patterns such as identifying bidders that tend to constantly being awarded contracts by the same buyers.

Public Procurement Contracts Database

Sources, Coverage and Structure

The procurement database covers 8 priority countries (Albania, Austria, Bulgaria, Czech Republic, Germany, Hungary, Italy, and Poland) selected to combine older and newer member states, and candidate countries in the Western Balkans, giving variation in institutional quality and data availability. The database follow the data model established with the DIGIWHIST project, which defines the common schema for harmonising heterogeneous procurement records that are comparable at contract-level data. The data collected for the NeT-RoL project follows these data standards and builds on already established criteria, such as improved bidder identifications and improved historical data. For each country, calls for tender and award notices were gathered from the national procurement source and from Tender Electronic Daily (TED).

Each country dataset contains a standardized set of variables to entail the necessary variables for linking it to different datasets (e.g. IDs) or cross-checking and verification with the collected notices. The specific national sources as well as the number of collected contracts from each source is listed in Table 1.

Table 1: Primary data sources by country

Country	Source	Contracts
AL	https://www.app.gov.al	287,355
AT	https://www.data.gv.at	93,304
	http://ted.europa.eu	34,709
	http://ted_eforms	10,889
	http://data.europa.eu/	6,019
BG	http://www.aop.bg//	181,656
	http://ted.europa.eu	82,586
	https://app.eop.bg	74,672
	http://ted_eforms	15,386
CZ	http://ted.europa.eu	187,196
	https://vestnikverejnychzakazek.cz	180,830
	http://ted_eforms	45,828
	https://vvz.nipez.cz	13,027
	http://data.europa.eu/	10,765
DE	http://ted.europa.eu	584,738
	http://ted_eforms	158,177
	http://data.europa.eu/	66,584
HU	http://kozbeszerzes.hu/	80,494

Country	Source	Contracts
	http://ted.europa.eu	64,644
	http://EKR	51,938
	http://kozbeszerzes.hu/csv	43,442
	http://ted_eforms	14,822
	http://data.europa.eu/	6,795
IT	https://dati.anticorruzione.it/opendata/download/dataset/	2,017,401
PL	ftp://ftp.uzp.gov.pl	2,538,501
	http://ted.europa.eu	1,269,367
	https://ezamowienia.gov.pl	588,411
	http://websrv.bzp.uzp.gov.pl	232,415
	http://ted_eforms	207,879
	http://data.europa.eu/	52,428

The data is organized at the lowest level observable per country (e.g. bid or lot) whereby each row corresponds to a bid submitted for a lot within a tender. This relates to the finest grain at which an award outcomes and prices are recorded. Higher levels of variable exploration can be aggregated through stable identifiers (see Table A1), such as lot or tender identification, or buyer and bidder identifications. The developed structure preserves the single bid detection, multi-lot tenders or consortium bidding.

The key identifiers and structural fields include the tender, lot and bid identifiers; buyer and bidder identifiers (and/or names), procedure types, publication and decision dates, price-related information, CPV and NUTS codes. Moreover, based on the processed data and through the DIGIWHIST pipeline the uploaded datasets contain the integrity indicators for the relevant red flags (e.g. single bidding, advertisement period, non-open procedure). The full variable list and definition is provided in Annex, Table A1.

Although data availability for some countries goes back to 2007, a more stable and better data quality starts after 2010/2011, and it particularly improves after the 2015 (with the exception of Austria whose data quality and size only increases from 2020 onwards).[4] The lower level of observable contracts per country (e.g. Germany) is accounted as a result of the national thresholds for data publication.

[4] After this period we can observe better buyer and bidder identifiers, more complete information provided and collected in tendering notices, decrease of missing information.

Table 2: Number of contracts per country-year

Year	AL	AT	BG	CZ	DE	HU	IT	PL	Total
2006	0	803	0	5,350	9,908	1,233	0	7,538	**24,832**
2007	0	1,051	8,959	6,723	12,243	1,253	0	9,499	**39,728**
2008	0	1,389	13,788	8,113	12,543	1,433	0	108,612	**145,878**
2009	0	1,553	10,887	11,041	15,210	1,504	0	289,075	**329,270**
2010	3,590	1,223	9,772	11,294	16,680	1,372	0	329,615	**373,546**
2011	4,058	2,008	16,727	11,135	25,715	1,384	55,140	421,335	**537,502**
2012	3,739	2,022	20,959	15,555	25,781	8,273	67,172	430,555	**574,056**
2013	20,613	2,010	19,864	21,580	27,701	19,589	107,305	434,427	**653,089**
2014	35,406	2,112	21,693	23,288	28,905	20,814	92,417	396,790	**621,425**
2015	23,806	2,185	19,022	22,703	31,613	19,881	133,340	332,210	**584,760**
2016	21,693	2,296	17,987	24,088	37,432	14,604	99,973	244,669	**462,742**
2017	20,002	2,311	20,106	33,843	43,770	16,367	118,581	243,208	**498,188**
2018	22,539	2,511	20,415	40,359	49,174	19,802	140,610	161,168	**456,578**
2019	19,842	7,338	21,476	39,535	57,461	15,627	121,291	156,226	**438,796**
2020	17,497	14,435	23,481	39,182	63,374	13,387	111,111	155,526	**437,993**
2021	16,867	17,782	31,726	31,146	66,940	13,867	138,249	188,376	**504,953**
2022	18,805	17,485	27,156	32,024	67,729	12,964	172,706	236,205	**585,074**
2023	18,967	19,672	28,887	28,062	68,992	13,367	194,530	248,359	**620,836**
2024	20,467	22,597	10,903	19,647	72,340	12,383	311,384	242,447	**712,168**
2025	19,464	22,138	10,492	12,978	75,988	9,555	153,592	253,161	**557,368**
Total	**287,355**	**144,921**	**354,300**	**437,646**	**809,499**	**218,659**	**2,017,401**	**4,889,001**	**9,158,782**

Table 3 reports the data overview per country in terms of number of contract-level observations (bids) and tenders. The buyers and suppliers' columns indicate the unique number of buyers and suppliers available in each dataset and based on their identifications. The last column (value) represents the total value of all contracts within each dataset. Poland has the highest number of available contracts, while Italy's total value is the highest. Albania, as a candidate country lags behind the EU countries. Beyond the information presented in Table 3, the uploaded datasets contain much richer and detailed information regarding procedure types, number of received bids, sectoral or regional breakdowns.

Table 3: Key descriptives per country

Country	Bids	Tenders	Buyers	Suppliers	Years	Value (€bn)
Albania	287,355	287,355	2,381	9,794	2010–2025	11.3
Austria	144,921	124,201	2,871	17,474	2006–2025	103.9
Bulgaria	354,300	169,810	4,264	26,883	2007–2025	78.9
Czech Rep.	437,646	214,296	13,059	26,839	2006–2025	289.8
Germany	809,499	524,962	15,023	63,955	2006–2025	515.2
Hungary	262,135	164,360	14,778	32,473	2006–2025	133.6
Italy	2,017,401	1,678,311	24,210	247,394	2011–2025	1,852.7
Poland	4,889,001	2,262,446	57,269	148,479	2006–2025	720.2
Total	**9,202,258**	**5,425,741**	**133,855**	**573,291**	**2006–2025**	**3,705.7**

Court decisions and debarment lists

In addition to public procurement data, within task 6 (subtask 6.2.2), publicly available sanctions and/or debarred companies (blacklists) have also been collected and processed.[5] Linking these records to the contract data makes it possible to identify economic operators that have been sanctioned, found in breach or formally excluded from participating in public procurement tendering procedures. All sources are official government or judicial publications. Because the collected records or available information varies across countries an initial mapping and a standardized template has been developed for either a judicial/arbitration decision (sanction for contract-breach or dispute) or administrative debarment (blacklist or formal exclusion of companies from participating for a defined period).

[5] This technical report and datasets include only a subset of merged public procurement and sanctions/blacklisted data.

The public procurement datasets integrate a sanction component from judicial decisions or blacklisted component from debarment lists. Sanctions and/or blacklisted data have been collected and merged into the public procurement data for the Czech Republic, Albania, Austria and Hungary. Data for Bulgaria is also collected. These records attach to the economic operators (bidders/suppliers) whether the company has been sanctioned, how many sanctions or judicial processes have been noted in the dataset against the company, the earliest and latest available decisions (Table 4).

Table 4: Sanctions integration into PP data

Procurement record			Sanctions attributes (joined on Bidder ID)				
Tender	Buyer	Bidder	Sanction flag	Cases	First case	Last case	Sanction ID
T-1001	Buyer A	Bidder 1	Yes	3	2014	2019	Includes all present sanctions
T-1002	Buyer A	Bidder 2	No	—	—	—	Includes all present sanctions
T-1003	Buyer B	Bidder 1	Yes	10	2014	2019	Includes all present sanctions
T-1004	Buyer B	Bidder 3	Yes	1	2021	2021	Includes all present sanctions
T-1005	Buyer C	Bidder 2	No	—	—	—	Includes all present sanctions
T-1006	Buyer C	Bidder 4	No	—	—	—	Includes all present sanctions
T-1007	Buyer C	Bidder 3	Yes	1	2021	2021	Includes all present sanctions

The collection is being carried out country by country and merged into the procurement data as each national source is validated. The current status of collection and integration is presented in Table 5. Judicial decisions data (sanctions) have been collected and merged for the Czech Republic, Albania, and Austria, while administrative data have been collected and merged for Hungary. Additionally, EU or Global-wide lists have also been collected and can be subset for the priority countries. Work on collecting decisions and blacklists for the remaining countries is in progress.

The linking of sanctions and/or blacklists with public procurement data is performed through the company identifiers. Prior to merging both datasets, a bidder id standardization is performed to match the patterns. Since a single company can make several decisions, i.e., there have been multiple processes against it, the sanctions lists are first consolidated at the company level so that contract data are not inadvertently multiplied. Each company receives a binary indicator of whether it has been sanctioned (yes/no), a count of distinctive cases that are included in the sanction's dataset, as well as the dates of the earliest and last case. Similarly, debarred companies are coded with a binary blacklisting indicator. Where applicable, this is accompanied by the corresponding debarment period, recorded as a from-to data range. This structure allows the debarment

status to be merged and cross-checked with the procurement data at a specific point in time.

Table 5: Sanctions and blacklists data collection and integration

Country	Decision type	Match key	Status
Czech Rep.	Court / arbitration decisions	Company identifier	Collected and merged
Albania	Court decisions	Company identifier	Collected and merged
Austria	Court decisions	Company identifier	Collected and merged
Hungary	Debarment / blacklist	Company identifier	Collected and merged
Bulgaria	Court / debarment	—	In progress
Germany	Court / debarment	—	Collected
Italy	Court / debarment	—	In progress
N. Macedonia	Court / debarment	—	Collected
Poland	Court / debarment	—	In progress

Network Representation

The procurement data are also processed and represented as bipartite supplier-buyer contracting networks. In each network there are two types of node: contracting authorities (buyers) and economic operators (suppliers/bidders), and an edge is drawn between a buyer and a supplier whenever a supplier was awarded a contract by that buyer. To keep the two node sets disjoint, both buyers and suppliers are prefixed distinctly in the generated dataset ("B_" for buyers and "S_" for suppliers).

Each edge carries two weights: the number of distinct contracts awarded between the buyer and supplier, as well as the total value of the awarded contracts (in comparable EU terms). From the bipartite network graph, a one-mode projection can be derived onto the buyer set and onto the supplier set in order to capture buyers that share suppliers and suppliers that share buyers (Table 6). For each country the network layer comprises of an edgelist[6] table consisting of buyer node, supplier node, number of contracts, and value (see Table A2 in the Annex).

[6] An edgelist represents a data structure in which each row represents a single connection (edge) between two entities (in our case a buyer and a bidder). The connection is represented through the awarded contract.

Table 6: Network representation illustration

Nodes (bipartite edge)		Edge weights	
Buyer node	Supplier node	Contracts	Total value (€)
B_100	S_100	12	4,850,000
B_100	S_101	3	920,000
B_100	S_102	1	145,000
B_101	S_100	7	2,310,000
B_101	S_103	5	1,680,000
B_102	S_101	9	3,275,000
B_102	S_104	2	560,000
B_103	S_100	4	1,120,000

Data Quality and Access

The datasets are collected, processed, and constructed in a scripted and version-conrtrolled pipelines so that every step is reproducible and can be traced from the raw published source to the published dataset and table. The merging between public procurement dataset and sanctions or blacklists respectively is additionally manually validated through a selected random sample that checks whether the collected sanction information can be traced to the original published web page and the collected information is reliable and precise.

The datasets are published on Mendeley (see link in footnote)[7]. The datasets are grouped based on the available information: public procurement, public procurement with sanctions and/or blacklists, and network.

[7] Mendeley link: https://data.mendeley.com/datasets/nfhtfwn5nk/1

References

Fazekas, Mihály, and Gábor Kocsis. 2020. "Uncovering high-level corruption: cross-national objective corruption risk indicators using public procurement data." B*ritish Journal of Political Science* 50, no. 1: 155-164.

Fazekas, Mihály, István János Tóth, and Lawrence Peter King. 2016. "An objective corruption risk index using public procurement data." European Journal on Criminal Policy and Research 22: 369- 397.

Fazekas, Mihály; Bence, Tóth; Abdou, Aly & Ahmed Al-Shaibani (2024) Global Contract-level Public Procurement Dataset. Data in Brief, 54.

Government Transparency Institute (2025). *Public procurement data processing* (Version 1.0). GTI-R/2025:06. Budapest: Government Transparency Institute.

Opentender portal: DIGIWHIST (2018). *Opentender.eu*. Horizon 2020 (Grant Agreement No. 645852). https://opentender.eu/

Annex

Variable group	**Variable**	**Definition**
ID variables	Buyer ID	Unique identifier of the contracting authority (buyer).
	Buyer name	Name of the contracting authority.
	Bidder ID	Unique identifier of the economic operator (bidder).
	Bidder name	Name of the economic operator/bidder.
	Source	Originating data source of the record (national portal, TED, etc.).
OpenTender Red Flags	Single bid	Integrity indicator flagging contracts awarded on a single received bid.
	Advertisement period	Integrity indicator based on the length of the advertisement period.
	Decision period	Integrity indicator based on the length of the decision period.
	Call for tender publication	Integrity indicator on whether the call for tender was published.
	Procedure type	Integrity indicator based on the procurement procedure used.
	Tax haven	Integrity indicator flagging suppliers registered in tax havens.
	Benford's law (bid prices)	Integrity indicator testing bid prices against Benford's law.
	Distinct market	Integrity indicator based on the distinctiveness of the market.
	Supplier share in buyer spending	Integrity indicator measuring a supplier's share of a buyer's spending.
	Contract modification	Integrity indicator flagging post-award contract modifications.
Red-flag input variables	Procedure type	Procurement procedure type (standardized classification).
	National procedure type	Procedure type as recorded under national classification.
	Bids count	Total number of bids received for the lot.
	Valid bids count	Number of valid bids received for the lot.
	First call for tender date	Date of the first call-for-tender publication.
	Last call for tender date	Date of the last call-for-tender publication.
	Bid deadline	Deadline for the submission of bids.
	Award decision date	Date on which the award decision was taken.
	Contract signature date	Date on which the contract was signed.
	Award period length	Number of days between the call for tender and the award decision.
	Bidder country	Country of registration of the economic operator.
	Bid price	Price of the bid (as recorded).
	Bid price (original)	Bid price in the original currency of the notice.
	Bid price (EUR)	Bid price converted to euros (DIGIWHIST conversion).
	Consortium bid	Indicator of whether the bid was submitted by a consortium.
	Award criteria count	Number of award criteria applied in the tender.
Control-	Main CPV	Main Common Procurement Vocabulary (CPV) code of the tender.

related variables	CPV codes	Full set of CPV codes associated with the tender.
	Buyer NUTS (robust)	Buyer region (NUTS), robust/cleaned version.
	Buyer NUTS	Buyer region under the NUTS classification.
	Bidder NUTS (robust)	Bidder region (NUTS), robust/cleaned version.
	Bidder NUTS	Bidder region under the NUTS classification.
	Buyer country	Country of the contracting authority.
	Bidder country	Country of the economic operator.
	Tender year	Year of the tender.
	Supply type	Type of supply (works, goods, or services).
	Selection method	Method used to select the winning bid (MEAT/Lowest price).
	Buyer type	Type of contracting authority (e.g. national, regional, utility).
	Framework agreement	Indicator of whether the tender is a framework agreement.
	Tender price (EUR)	Tender value converted to euros (DIGIWHIST conversion).
	Bid price (EUR)	Bid value converted to euros (DIGIWHIST conversion).
	Estimated tender price (PPP)	Estimated tender value in purchasing-power-parity terms.
	Bid price (PPP)	Bid value in purchasing-power-parity terms.
Sanctions-related variables	Sanction flag	Whether the economic operator appears in the court/sanctions register (Yes/No)
	N cases	Number of cases recorded against the bidder
	First year case	Year of the bidder's earliest case (from the decision date)
	Last year case	Year of the bidder's most recent case (from the decision date)
	Filing date	Most recent non-missing filing date across the bidder's cases
	Decision date	Most recent non-missing decision date across the bidder's cases
	Raw date legal force	Most recent non-missing date on which a decision entered into legal force, across the bidder's cases
	Sanction ID	Stable identifier of the linked decision(s). Multiple cases are concatenated with (";")
Blacklist-related variables	Active blacklist flag	Whether the contract was awarded while the bidder was under an active exclusion (Yes/No)
	Ever blacklisted	Whether the bidder appears in the debarment register at all, regardless of whether any exclusion was active at award time (Yes/No)
	Blacklist effective start	Start date of the relevant debarment: in force at award time where one applies, otherwise the firm's most recent recorded exclusion
	Blacklist effective end	End date of the relevant debarment
	Blacklist decision date	Date of the administrative decision for debarment period
	Blacklist decision reference	Official reference of the debarment decision
	Sanction ID	Stable identifier of the linked debarment record

Table A2: Variables of edgelist dataset – network representation

Field	**Type**	**Description**
Buyer node	string	Contracting authority identifier, prefixed B_
Bidder node	string	Economic operator identifier, prefixed S_
N contracts	integer	Number of distinct contracts linking the pair
Value (EUR)	numeric	Total contract value between the pair, EUR